

nature



HIGH FIDELITY

Silicon qubits cross key error-correction threshold for quantum computing

Coronavirus

Counting the global death toll of the COVID-19 pandemic

Pregnancy health

Cell-free RNA in blood samples can predict risk of pre-eclampsia

Precision dating

Detailed timelines for global networks in the Viking Age

Nature.2022.01.22

[Sat, 22 Jan 2022]

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

This Week

- **[The ‘war on cancer’ isn’t yet won](#)** [19 January 2022]
Editorial • The US National Cancer Act of 1971 has fostered tremendous progress in our understanding of the biology that underlies cancer. However, scientific and social challenges remain.
- **[Biodiversity faces its make-or-break year, and research will be key](#)** [19 January 2022]
Editorial • A new action plan to halt biodiversity loss needs scientific specialists to work with those who study how governments function.
- **[Message to mayors: cities need nature](#)** [17 January 2022]
World View • Urban areas must invest in ecosystems, not just in roads, pipes and homes.
- **[Slime moulds souped up with plant genes churn out drugs](#)**
[17 January 2022]
Research Highlight • The amoeba life stage of a common slime mould outshines workhorse microbes at producing certain medicinal compounds.
- **[Synthetic viral lookalikes sneak gene-editing tools into cells](#)**
[13 January 2022]
Research Highlight • Specially engineered particles allow researchers to fix blindness-causing genes in mice.
- **[Drug-fuelled parties helped ancient Andean rulers to hold power](#)** [12 January 2022]
Research Highlight • Guests quaffed beer infused with hallucinogenic seeds, according to the detritus from a feast in what is now Peru.
- **[Humans and bacteria rely on the same aeons-old immune defence](#)** [13 January 2022]
Research Highlight • Bacteria fending off viruses rely on a form of cell death also used by mammals to fight infection.
- **[Vast fish breeding colony is more than twice the size of Paris](#)** [13 January 2022]
Research Highlight • Roughly 60 million nests of the Jonah’s icefish cover 240 square kilometres — the largest unbroken fish breeding ground ever found.

- **Print job completed: a bendable image display** [07 January 2022]

Research Highlight • A display based on organic light-emitting diodes has now been produced entirely by 3D printing.

- **Cosmic crash explains a mystery on the Moon** [11 January 2022]

Research Highlight • Simulations show how a violent impact long ago might have led to lopsided lunar chemistry.

- EDITORIAL
- 19 January 2022

The ‘war on cancer’ isn’t yet won

The US National Cancer Act of 1971 has fostered tremendous progress in our understanding of the biology that underlies cancer. However, scientific and social challenges remain.



Richard Nixon signed the National Cancer Act at the White House on 23 December 1971. Since then, the idea of an ‘all-out assault’ on cancer has been moderated, with progress understood to be more likely to occur in small steps. Credit: Linda Bartlett/NIH/National Cancer Institute

In 1970, advisers to the US Senate attempted to forecast the evolution of cancer treatment. “The long-term future may belong to the immunologist and the geneticist, the intermediate future to the chemotherapist,” they

wrote, “but the present and the immediate future belong in the main to the surgeon and to some extent to the radiologist.” More than 50 years later — and well into the ‘long-term future’ of genetics and immunology — their predictions have proved remarkably prescient.

The advisers had been called in to consult on what would become the National Cancer Act of 1971, the opening salvo of then-president Richard Nixon’s “war on cancer”. Cancer was the second leading cause of death in the United States, and the act — passed in December 1971 — called for US\$1.5 billion to be poured into cancer research over three years, equivalent to more than \$10 billion today.

The act strengthened the nation’s cancer-research infrastructure, creating a national clinical-trials network, as well as a host of specialized cancer research centres. But, from the start, Nixon’s focus on finding a swift cure for cancer was criticized as naive. Some researchers cautioned that cancer is a complex disease, and that progress was likely to be made in small, hard-won steps rather than a single, decisive cure. That prediction also turned out to be correct.

Half a century after the National Cancer Act became law, cancer is still the second leading cause of death in the United States. The act has fostered tremendous advances, in large part through a greater understanding of the biology that underlies the disease. But cancer research still faces challenges, not all of them scientific. Many therapies are too expensive for individuals or health-care systems: in the United States, 42% of people with cancer experienced severe financial hardship within two years of diagnosis, and, in many countries, innovative cancer therapies such as immunotherapies are out of reach for the majority.

In 1971, there were already signposts indicating the directions in which cancer research could lead. It was the year in which researchers published results of the first clinical trial of tamoxifen, a hormonal therapy for breast cancer. The US Food and Drug Administration (FDA) approved the drug, in 1978, to treat advanced cancers, and tamoxifen has since become a mainstay of breast cancer therapy.

When the act was passed, researchers already knew that the genomes of cancerous cells were chaotic and unstable. By 1976, cancer cytogeneticist Peter Nowell had realized that such genetic variability would make the task of finding therapies that much harder and more complex. Today, we know that cancer is not one but many different diseases. The complexities are mind-boggling — which is one of the reasons why cancer is so difficult to cure. Large-scale tumour-sequencing projects have taught us more about how each cancer type can be broken down into subtypes of subtypes, and that each individual tumour has a unique molecular make-up. Furthermore, this molecular signature shifts as the cancer progresses and responds to treatment.

Over the past 50 years, scientists have made remarkable progress in understanding this complexity through projects carried out by basic and clinical researchers working ever-more closely. The Cancer Genome Atlas and Pan-Cancer Analysis of Whole Genomes projects have revealed both the complexity of cancer and important clues to new treatments. Discoveries made during basic research on the cell cycle have led to the development of anticancer drugs that target certain cell-cycle proteins.

Last year, the FDA approved a drug for some non-small cell lung cancers that disables a mutant form of the KRAS protein. The *KRAS* gene is closely related to the first human oncogene — a gene with the potential to cause cancer — to be discovered, a decade or so after the act was passed. And therapies that target the immune system have opened up new avenues of treatment for several cancers.

Such progress has saved lives. In 1971, only half of people in the United States diagnosed with cancer would live for more than five years beyond that diagnosis; now the fraction is around two-thirds. Scientists have since recognized the value of early detection, and in many countries there are early-detection programmes for some of the most common cancers, including breast and colon cancers.

The advances fuelled by the National Cancer Act have also spilt over into other fields. The study of how cancer cells interact with one another in tumours has fuelled a broader understanding of cell–cell interactions. Synthetic immune cells designed to fight cancer are now being studied as

potential treatments for conditions including fibrosis of the heart and lungs. And large clinical trials designed to test multiple interventions and to adapt to interim data — a novel study design pioneered in cancer research — have been pivotal to the discovery of ways to treat COVID-19.

Today, scientists rarely talk of a broad cure for cancer. Instead, as Norman Sharpless, the director of the US National Cancer Institute, says, the aim is to end cancer “as we know it”. Some cancer might be an inevitable consequence of ageing, he says, but researchers can make headway against the shortcomings of current cancer therapies, by, for example, tackling the cancers for which there is little to offer patients; finding ways to reduce the side effects of harsh cancer therapies; and addressing the racial and socio-economic inequalities that affect access to clinical trials and treatments. The bulk of the data that have been painstakingly gathered over the past half-century were collected from white people — a bias that must be corrected.

The combination of technological advances with continued collaboration between basic and clinical researchers could sustain the momentum generated by the act into the next 50 years — and take the field ever further from the days when surgery and radiotherapy were the only treatments.

Nature **601**, 297 (2022)

doi: <https://doi.org/10.1038/d41586-022-00109-3>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00109-3>

- EDITORIAL
- 19 January 2022

Biodiversity faces its make-or-break year, and research will be key

A new action plan to halt biodiversity loss needs scientific specialists to work with those who study how governments function.



Targeted measures can help to stop extinctions, including of Père David's deer (*Elaphurus davidianus*), but conserving biodiversity will also require combating climate change, cutting pollution and enhancing sustainable food systems. Credit: Staffan Widstrand/Wild Wonders of China/Nature Picture Library

Biodiversity is being lost at a rate not seen since the last mass extinction. But the United Nations decade-old plan to slow down and eventually stop the decline of species and ecosystems by 2020 has failed. Most of the plan's 20 targets — known as the Aichi Biodiversity Targets — [have not been met](#).

The Aichi targets are part of an international agreement called the UN Convention on Biological Diversity, and member states are now finalizing replacements for them. Currently referred to as the post-2020 global biodiversity framework (GBF), the new targets are expected to be agreed this summer at the second part of the convention's Conference of the Parties (COP15) in Kunming, China. The meeting was due to be held in May, but is likely to be delayed by a few months. Finalizing the framework will be down to government representatives working with the world's leading biodiversity specialists. But input from social-science researchers, especially those who study how organizations and governments work, would improve its chances of success.

A draft of the GBF was published last July. It aims to slow down the rate of biodiversity loss by 2030. And by 2050, biodiversity will be “valued, conserved, restored and wisely used, maintaining ecosystem services, sustaining a healthy planet and delivering benefits essential for all people”. The plan comprises 4 broad goals and 21 associated targets. The headline targets include conserving 30% of land and sea areas by 2030, and reducing government subsidies that harm biodiversity by US\$500 billion per year. Overall, the goals and targets are designed to tackle each of the main contributors to biodiversity loss, which include agriculture and food systems, climate change, invasive species, pollution and unsustainable production and consumption.

The biodiversity convention's science advisory body is reviewing the GBF and helping governments to decide [how the targets are to be monitored](#). But researchers and policymakers have been writing biodiversity action plans since the 1990s, and most of these strategies have failed to make a lasting impact on two of the three key demands: that global biodiversity be conserved and that natural resources be used sustainably.

Some of these failures are to do with governance, which is why it is important to involve not just researchers in the biological sciences, but also

people who study organizations and how governments work. This knowledge, when allied to conservation science, will help policymakers to obtain a fuller picture of both the science gaps and the organizational challenges in implementing biodiversity plans.

The GBF is a comprehensive plan. But success will require systemic change across public policy. That is both a strength and a weakness. If systemic change can be implemented, it will lead to real change. But if it cannot, there's no plan B. This has led some researchers to argue that one target or number should be prioritized, and defined in a way that is clear to the public and to policymakers. It would be biodiversity's equivalent of the 2 °C climate target. The researchers' "rallying point for policy action and agreements" is to keep species extinction to well below 20 per year across all major groups ([M. D. A. Rounsevell et al. *Science* 368, 1193–1195; 2020](#)). Such focus does yield results. A study published in *Conservation Letters* found a high probability that targeted action has prevented 21–32 bird and 7–16 mammal extinctions since 1993 ([F. C. Bolam et al. *Conserv. Lett.* 14, e12762; 2021](#)). Extinction rates would have been around three to four times greater without conservation action, the researchers found.

But not all agree that just one target should be given priority. A group of more than 50 biodiversity researchers from 23 countries point out in a policy report this week (see [go.nature.com/3fv8oiv](#)) that data on species are distributed unequally: 10, mostly high-income, countries account for 82% of records.

The researchers also modelled how different scenarios would affect the GBF's 21 targets. They found that achieving the targets would require action in all of the target areas — not just a few. Focusing strongly on just one or two targets — such as expanding protected areas — will have, at best, a modest impact on achieving the UN convention's goals and targets.

The difficulty in getting governments to adopt such an integrated approach is that they (as well as non-governmental organizations and businesses) tend to tackle sustainability challenges piecemeal. Actions from last November's climate COP in Glasgow, UK, will be implemented separately from those decided at the biodiversity COP because, in most countries, different government departments deal with climate change and biodiversity.

The science advisers for the biodiversity convention will meet in Geneva, Switzerland, in March to finalize their advice. They are not advocating reform of how governments organize themselves to implement policies in sustainable development — partly (and rightly) because this is generally beyond their fields of expertise. But it's not too late to consult those with the relevant knowledge.

In the past, the UN has commissioned social scientists, for example in the UN Intellectual History Project, a series of 17 studies summarizing the experience of UN agencies spanning gender equality, diplomacy, development, trade and official statistics. However, this work, which ended in 2010, did not assess what has and hasn't worked in science and environmental policy. Unless these perspectives are incorporated into biodiversity-research advice, any future plans risk going the way of their predecessors.

Nature **601**, 298 (2022)

doi: <https://doi.org/10.1038/d41586-022-00110-w>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00110-w>

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

- WORLD VIEW
- 17 January 2022

Message to mayors: cities need nature



Urban areas must invest in ecosystems, not just in roads, pipes and homes.

- [Akanksha Khatri](#) 0

My home town of New Delhi is battling with air pollution, contaminated water supplies and heatwaves. Just last November, schools were shut for more than a week because of untenable air quality. In the past 5 years, students have missed at least 26 days of school for this reason. The city is one of an increasing number of places where rapid, unplanned urbanization has undermined the environmental services essential for a good quality of life. The built environment grew globally by two-thirds between 2000 and 2012. Almost half of cities' economic output depends on natural systems that are threatened by climate change and biodiversity loss. By 2050, 70%

of people will live in cities; 3 in 4 of these urbanites will face water scarcity.

To thrive, cities must lean into nature. That means having open green spaces and interconnected waterways to prevent floods; green roofs and walls that reduce temperatures and produce food; and forests planted as green belts to oxygenate urban areas and regenerate ecosystems.

But cities are doing little of this. By one estimate, only 0.3% of spending on urban infrastructure goes towards ‘nature-based solutions’, efforts that use ecosystems to help mitigate pollution, reduce risks from floods and storms and provide healthy air, water, food and living conditions.

For the past four years, I’ve led efforts at the World Economic Forum in Cologny, Switzerland, to encourage businesses and governments to speed up the transition to what I call a nature-positive economy — to make the planet more resilient by reducing emissions and enriching ecosystem services. One initiative, ‘BiodiverCities by 2030’, helps city administrators to explicitly include existing ecosystems in urban planning and to keep economic, sociological and ecological needs in harmony. This week, we published a report on how urban leaders can use nature-based solutions in their planning (go.nature.com/3fzhqss).

Historically, cities developed where nature could meet human needs: near water, on fertile land, where conditions were clement. Most cities have forgotten those roots: the physical infrastructure and social institutions that make cities what they are increasingly impair the very conditions that enabled cities to thrive. The risk of floods, storms or fires is growing because of climate change or degraded ecosystems. But most project evaluations do not consider these risks or ways to mitigate them.

Models and research analyses allow city planners to rigorously compare return on investments from ‘grey infrastructure’, based on cement, with those from asphalt and green infrastructure that harness ecosystem services. City planning usually defaults to the former, but this is more costly than is assumed. Last year, the International Institute for Sustainable Development calculated that nature-based infrastructure could contribute to climate resilience, irrigation, sanitation and more at 50% of the cost of conventional

built infrastructure. This is largely because of lower capital investments and costs incurred owing to extreme weather.

One goal of BiodiverCities by 2030 is to help mayors access the research and analyses that are most relevant for their cities. Over the past year, we have brought together political and business leaders and specialists on urban development, conservation, bioeconomics and more. It has become clear how unfamiliar most city leaders are with the opportunities for expanding nature in the built environment, and this finding informed our report.

There are examples of successful projects across the globe. Lena Chan, senior director of the biodiversity conservation division of the National Parks Board of Singapore, and co-chair of the Global Commission on BiodiverCities by 2030, helped five ministers to plan for and implement urban renaturing in the city-state. Among other improvements, the result was vastly expanded green spaces — along roadsides, on rooftops and even on the sides of buildings. This will boost quality of life by providing cooler urban temperatures and resilience against sea-level rise and inland flooding. The mayor of Sierra Leone's capital has launched a campaign, 'Freetown the Treetown', to nurse 15 tree species over 11 sites; this will reduce the risk of floods and landslides. By the end of 2022, the area covered by vegetation will have expanded by 50%.

Such projects require planning on many levels, not least regarding equity. Tree-planting and landscaping efforts can lead to 'green gentrification', with residents on lower incomes being priced out of restored neighbourhoods if local socio-economic considerations are ignored. Good data help. A 2021 study found that apartment blocks in low-income areas of US cities had 15% less tree cover and were 1.5 °C hotter than those in high-income areas ([R. I. McDonald et al. PLoS ONE 16, e0249715; 2021](#)). Citizen science can help in democratizing efforts and promoting necessary changes. For example, BiodiverCities by 2030 crowdsourced international projects in which people monitored biodiversity around their homes using mobile-phone apps.

Cities can protect themselves by shoring up natural ecosystems. Last October, the Indian state of Kerala was hit by heavy floods. The city of

Kochi was particularly badly affected, but, as Kochi's mayor told me, the disaster could have been much worse. In 2020, when the Kochi Municipal Corporation launched plans to restore canals, it included the restoration of wetlands and planting of mangroves.

But much more must be done. To achieve resilience, ecosystems must be part of urban infrastructure.

Nature **601**, 299 (2022)

doi: <https://doi.org/10.1038/d41586-022-00102-w>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00102-w>

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

- RESEARCH HIGHLIGHT
- 17 January 2022

Slime moulds souped up with plant genes churn out drugs

The amoeba life stage of a common slime mould outshines workhorse microbes at producing certain medicinal compounds.



The slime mould *Dictyostelium discoideum* has a vegetative form (above) that can turn into an amoeba form useful for making drugs. Credit: B. Kelly *et al./Nature* ([CC BY 4.0](#))

The amoeba form of a slime mould can be engineered to produce a range of natural, medically useful compounds, including several that are used to treat nerve damage and pain¹.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Nature **601**, 300 (2022)

doi: <https://doi.org/10.1038/d41586-022-00067-w>

References

1. 1.

Reimer, C. *et al.* *Nature Biotechnol.* <https://doi.org/10.1038/s41587-021-01143-8> (2022).

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

- RESEARCH HIGHLIGHT
- 13 January 2022

Synthetic viral lookalikes sneak gene-editing tools into cells

Specially engineered particles allow researchers to fix blindness-causing genes in mice.



A type of genome-editing tool called a base editor (bound to DNA; artist's illustration) can be smuggled into an animal's cells by a virus-like particle. Credit: Ramon Andrade 3DCIENCIA/Science Photo Library

Researchers have engineered virus-like particles that efficiently deliver gene-editing proteins to specific organs in living animals¹.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 300 (2022)

doi: <https://doi.org/10.1038/d41586-022-00068-9>

References

1. 1.

Banskota, S. *et al.* *Cell* <https://doi.org/10.1016/j.cell.2021.12.021> (2022).

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

- RESEARCH HIGHLIGHT
- 12 January 2022

Drug-fuelled parties helped ancient Andean rulers to hold power

Guests quaffed beer infused with hallucinogenic seeds, according to the detritus from a feast in what is now Peru.



Beer for a feast might have been served to guests in this cup excavated in Quilcapampa, an outpost of the ancient Wari empire in the Andes. Credit: Luis Manuel González La Rosa, Justin Jennings/Royal Ontario Museum

Centuries-old plant remains suggest that beer spiked with a hallucinogenic drug played a crucial part in helping an ancient superpower to maintain political control over the Andes¹.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 300 (2022)

doi: <https://doi.org/10.1038/d41586-022-00066-x>

References

1. 1.

Biwer, M. E., Yépez Álvarez, W., Bautista, S. L. & Jennings, J.
Antiquity <https://doi.org/10.15184/aqy.2021.177> (2022).

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

- RESEARCH HIGHLIGHT
- 13 January 2022

Humans and bacteria rely on the same aeons-old immune defence

Bacteria fending off viruses rely on a form of cell death also used by mammals to fight infection.

 Negative stain electron microscopy of Runella gasdermin pores in DOPC liposomes

Some bacteria have a molecular pathway that creates destructive holes in cell membranes (honeycomb structure at centre) in response to pathogens — a mechanism already known in mammals Credit: A. G. Johnson *et al./Science*

At first glance, bacteria and humans seem to have little in common. But scientists have discovered that the cells in our bodies and some of Earth's single-celled microorganisms use a similar strategy to fight off invaders¹.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 300 (2022)

doi: [*https://doi.org/10.1038/d41586-022-00005-w*](https://doi.org/10.1038/d41586-022-00005-w)

References

1. 1.

Johnson, A. G. *et al. Science*

[*http://www.science.org/doi/10.1126/science.abj8432*](http://www.science.org/doi/10.1126/science.abj8432) (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00005-w>

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

- RESEARCH HIGHLIGHT
- 13 January 2022

Vast fish breeding colony is more than twice the size of Paris

Roughly 60 million nests of the Jonah's icefish cover 240 square kilometres — the largest unbroken fish breeding ground ever found.



The shallow nests made by the Jonah's icefish speckle the floor of the Weddell Sea. Credit: Alfred Wegener Institute/PS124 OFOBS team

A truly massive breeding colony of fish has been discovered near Antarctica¹.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Nature **601**, 301 (2022)

doi: <https://doi.org/10.1038/d41586-022-00065-y>

References

1. 1.

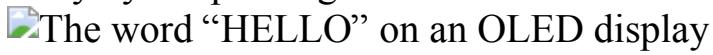
Purser, A. *et al.* *Curr. Biol.* <https://doi.org/10.1016/j.cub.2021.12.022> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00065-y>.

- RESEARCH HIGHLIGHT
- 07 January 2022

Print job completed: a bendable image display

A display based on organic light-emitting diodes has now been produced entirely by 3D printing.



A 64-pixel array produced by 3D printing can spell out a word by displaying one letter at a time (above, composite image). Credit: Ruitao Su ([CC BY 4.0](#))

Devices called organic light-emitting diodes (OLEDs) can be used to make computer screens and smartphone displays that offer better image quality and energy efficiency than other display technologies, such as liquid-crystal displays. Now, flexible OLED displays have been produced entirely by 3D printing¹.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 301 (2022)

doi: <https://doi.org/10.1038/d41586-022-00043-4>

References

1. 1.

Su, R., Park, S. H., Ouyang, X., Ahn, S. I. & McAlpine, M. C. *Sci. Adv.* <https://doi.org/10.1126/sciadv.abl8798> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00043-4>

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

- RESEARCH HIGHLIGHT
- 11 January 2022

Cosmic crash explains a mystery on the Moon

Simulations show how a violent impact long ago might have led to lopsided lunar chemistry.



The South Pole–Aitken Basin (central blue area) marks the spot where a gargantuan cosmic object ploughed into the Moon. Credit: NASA/GSFC/Univ. Arizona

A space smash-up that happened more than 4 billion years ago could explain one of the Moon’s biggest mysteries — why its nearside and far side look so different from one another¹.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 301 (2022)

doi: <https://doi.org/10.1038/d41586-022-00064-z>

References

1. 1.

Zhang, N. *et al.* *Nature Geosci.* <https://doi.org/10.1038/s41561-021-00872-4> (2022).

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

News in Focus

- **[Fossil-record bias and huge research database](#)** [19 January 2022]
News Round-Up • The latest science news, in brief.
- **[First pig-to-human heart transplant: what can scientists learn?](#)** [14 January 2022]
News • Researchers hope that a person who has so far lived for a week with a genetically modified pig heart will provide a trove of data on the possibilities of xenotransplantation.
- **[‘Killer’ immune cells still recognize Omicron variant](#)** [11 January 2022]
News • Amid concerns over lost antibody defences, some researchers argue that more attention should be paid to T cells.
- **[COVID vaccines safely protect pregnant people: the data are in](#)** [12 January 2022]
News • Despite evidence that pregnant people are at high risk of serious disease, many remain unvaccinated.
- **[How to protect US science from political meddling after Trump](#)** [11 January 2022]
News • In a fresh report, federal researchers recommend ways to strengthen scientific integrity and preserve public trust in government.
- **[Omicron thwarts some of the world’s most-used COVID vaccines](#)** [13 January 2022]
News • Inactivated-virus vaccines elicit few, if any, infection-blocking antibodies — but might still protect against severe disease.
- **[The pandemic’s true death toll: millions more than official counts](#)** [18 January 2022]
News Feature • Countries have reported some five million COVID-19 deaths in two years, but global excess deaths are estimated at double or even quadruple that figure.

- NEWS ROUND-UP
- 19 January 2022

Fossil-record bias and huge research database

The latest science news, in brief.



The fossilized bones of the dinosaur *Stegouros elengassen* were excavated in Chilean region of Patagonia, South America. Credit: Martin Bernetti/AFP/Getty

Fossil record skewed by rich countries

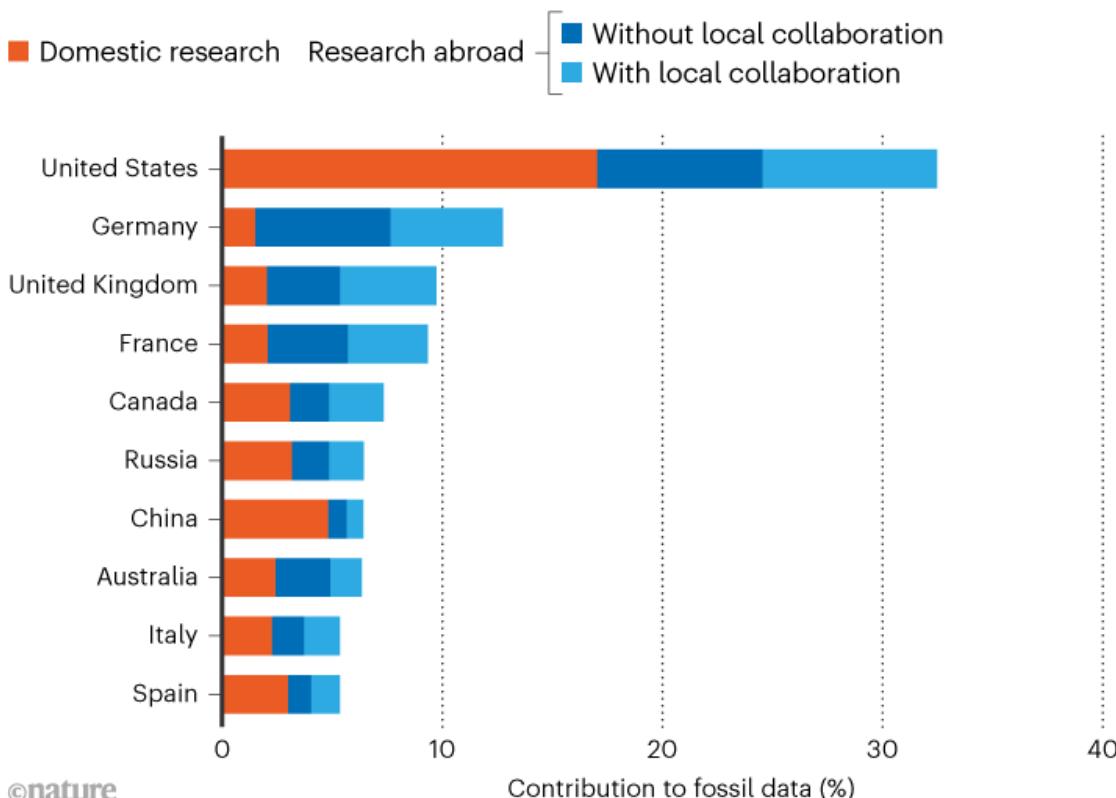
Our understanding of the history of life on Earth is [biased towards the views of wealthier countries](#), warns a study of the fossil record that found that a whopping 97% of palaeontological data come from scientists in high- and upper-middle-income countries ([N. B. Raja et al. *Nature Ecol. Evol.* <https://doi.org/hcrk; 2021>](#)).

“I knew it was going to be high, but I didn’t think it was going to be this high. It was astonishing,” says Nussäibah Raja, a palaeontologist at the Friedrich Alexander University of Erlangen–Nuremberg in Erlangen, Germany, who co-led the study. Raja and colleagues analysed data from the Paleobiology Database (PBDB), a widely used repository. The team examined the authorship of 29,039 papers indexed in the PBDB that were published between 1990 and 2020.

More than one-third of these records included authors based in the United States; the rest of the top five consisted of Germany, the United Kingdom, France and Canada (see ‘Global imbalance’).

GLOBAL IMBALANCE

The majority of palaeontological data come from researchers in high- and upper-middle-income countries.



Source: N. B. Raja *et al. Nature Ecol. Evol.* <https://doi.org/10.1038/s43246-021-01395-w> (2021)

The outsized influence of rich countries on palaeontology could lead to a warped view of life's history, the researchers say.

"Scientific knowledge should not be restricted to small parts of the planet," says Pedro Godoy, a palaeontologist at the Federal University of Paraná in Curitiba, Brazil. "Science certainly loses quality by being so restrictive."

Massive open scholarly index launches

A free index of hundreds of millions of scientific documents, which aims to provide an open catalogue of the global research system, has been launched.

The database, called OpenAlex after the ancient Library of Alexandria in Egypt, collates academic publications, author information, publication venues and research topics. Developed by scholarly services company OurResearch in Vancouver, Canada, the platform replaces Microsoft Academic Graph (MAG), a free alternative to subscription-based scholarly databases such as Scopus and Dimensions that was discontinued at the end of 2021.

OpenAlex draws its data from MAG's records and other sources, including ORCID and Crossref, says Jason Priem, co-founder of OurResearch. The idea is to interlink OpenAlex with several other databases to make it more comprehensive, he adds. "We now have much better coverage of open access than MAG ever did," Priem adds.

"It's just pulling lots of databases together in a clever way," says Euan Adie, founder of Overton, a London-based firm that tracks research cited in policy documents. Overton used to get data from different sources, including MAG, ORCID and Crossref, but it has now switched to using only OpenAlex, says Adie.

Nature **601**, 303 (2022)

doi: <https://doi.org/10.1038/d41586-022-00103-9>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00103-9>

- NEWS
- 14 January 2022

First pig-to-human heart transplant: what can scientists learn?

Researchers hope that a person who has so far lived for a week with a genetically modified pig heart will provide a trove of data on the possibilities of xenotransplantation.

- [Sara Reardon](#)



Surgeons at the University of Maryland Medical Center transplanted a genetically altered pig heart into David Bennett. Credit: University of Maryland School of Medicine

The first person to receive a transplanted heart from a genetically modified pig is doing well after the procedure last week in Baltimore, Maryland. Transplant surgeons hope the advance will enable them to give more people animal organs, but many ethical and technical hurdles remain.

“It’s been a long road to get to this point, and it’s very exciting we are at a point where a group was ready to try this,” says Megan Sykes, a surgeon and immunologist at Columbia University in New York City. “I think there’s going to be a lot of interesting things to be learned.”

Physicians and scientists worldwide have for decades been pursuing the goal of transplanting animal organs into people, known as xenotransplantation.

Unusual opportunity

Last week’s procedure marks the first time that a pig organ has been transplanted into a human who has a chance to survive and recover. In 2021, surgeons at New York University Langone Health transplanted kidneys from the same line of genetically modified pigs into two legally dead people with no discernible brain function. The organs were not rejected, and functioned normally while the deceased recipients were sustained on ventilators.

Aside from that, most research has so far taken place in non-human primates. But researchers hope that the 7 January operation will further kick-start clinical xenotransplantation and help to push it through myriad ethical and regulatory issues.

“From 4 patients, we’d learn a lot that we wouldn’t learn from 40 monkeys,” says David Cooper, a transplant surgeon at Massachusetts General Hospital in Boston. “It’s time we move into the clinic and see how these hearts and kidneys do in patients.”

Xenotransplantation has seen significant advances in recent years with the advent of CRISPR–Cas9 genome editing, which made it easier to create pig

organs that are less likely to be attacked by human immune systems. The latest transplant, performed at the University of Maryland Medical Center (UMMC), used organs from pigs with ten genetic modifications.

The researchers had applied to the US Food and Drug Administration (FDA) to do a clinical trial of the pig hearts in people, but were turned down. According to Muhammad Mohiuddin, the University of Maryland surgeon who leads the research team behind the transplant, the agency was concerned about ensuring that the pigs came from a medical-grade facility and wanted the researchers to transplant the hearts into ten baboons before moving on to people.

But 57-year-old David Bennett gave Mohiuddin's team a chance to jump straight to a human transplant. Bennett had been on cardiac support for almost two months and couldn't receive a mechanical heart pump because of an irregular heart beat. Neither could he receive a human transplant, because he had a history of not complying with doctors' treatment instructions. Given that he otherwise faced certain death, the researchers got permission from the FDA to give Bennett a pig heart.



The heart used in the transplant came from a pig with several genetic modifications, including some to knock out genes that trigger the human immune system. Credit: University of Maryland School of Medicine

The surgery went well and “the heart function looks great”, Mohiuddin says. He and his team will monitor Bennett’s immune responses and the performance of his heart. They will continue working towards controlled clinical trials, but Mohiuddin says they might apply to conduct more emergency procedures if the right patients come along.

If Bennett’s procedure proves successful and more teams try similar surgery, regulators and ethicists will need to define what makes a person eligible for a pig organ, says Jeremy Chapman, a retired transplant surgeon at the University of Sydney in Australia. Waiting a long time for an organ isn’t enough to justify the highly experimental and possibly risky procedure, he says. That’s especially true with other organs, such as kidneys; most people waiting for kidney transplants can be put on dialysis.

Chapman likens the process to the use of experimental cancer drugs that are too dangerous to test in people with other options. Regulators and ethicists will need to decide what chance of success outweighs the risk of making a person wait for a human organ, he says.

‘Crazy, exciting week’

For now, transplantation is limited by the supply of pigs as well as regulatory hurdles. There is currently just one company — Revivicor in Blacksburg, Virginia, owned by United Therapeutics — that has suitable facilities and clinical-grade pigs.

“This has been a crazy, exciting week,” says Revivicor chief executive David Ayares. The company’s pigs are being raised at a facility near Birmingham, Alabama, but Revivicor is building a larger facility in Virginia that it hopes will eventually supply hundreds of organs per year.

Ayares has been engineering pigs for two decades, testing how various genetic modifications limit rejection in humans and other primates. To make the pig heart used in the transplant, the company knocked out three pig

genes that trigger attacks from the human immune system, and added six human genes that help the body to accept the organ. A final modification aims to prevent the heart from responding to growth hormones, ensuring that organs from the 400-kilogram animals remain human-sized.

Although the combination seems to have worked, it's unclear how many of the modifications are necessary. "There's a lot more science needed in assessing each genetic modification," says Sykes, who adds that "we need that information" because the modifications also have the potential to be harmful in people.

Mohiuddin says that because each transplant into a baboon costs approximately US\$500,000, testing multiple combinations would be prohibitively expensive. Cooper and others say that the future of xenotransplantation probably includes tailoring the modifications to suit particular organs and recipients. Cooper's own research, for example, has found that in baboons that receive pig kidneys, the growth-hormone modification causes problems with urine transport. But he says that his team hopes to perform a kidney transplant into a person soon, if it can get a pig with the proper genetic modifications.

Whatever happens, it might be some time before other organs are ready for clinical use. Waiting lists are shorter for liver transplants, making it harder to justify people receiving a pig organ. And although people who require lung transplants often die on the waiting list, Sykes says that fragile pig lungs have proved tricky to transplant into primates and are often rejected.

Limits of animal models

Cooper, Chapman and others say that it is important to study the transplants in humans rather than baboons. The differences between species "preclude us moving further with that model to predict the clinical outcome", Chapman says.

Non-human primates tend to have antibodies that humans don't, which attack proteins on pig organs, so a lot of work has gone into making the organs suitable for baboons, not people.

Furthermore, researchers need to be able to study the pig heart's physiology — whether it will beat at the same rate as a human heart, for instance — and whether ill people will react to the transplant in the same way as healthy baboons.

Several other companies are engineering pigs for solid organ transplants with different genetic modifications, although none yet has medical-grade facilities, as United Therapeutics does. eGenesis in Cambridge, Massachusetts, is making pigs that cannot pass on retroviruses that are present in all pig genomes. And NZeno in Auckland, New Zealand, is breeding miniature pigs whose kidneys remain human-sized without growth-hormone modifications. Chapman suspects that many more organizations are genetically modifying pigs for transplant but have yet to disclose commercially sensitive information.

Ayares and United Therapeutics declined to say how much each pig costs to produce, although they acknowledge that the animals are expensive. But as more companies get into the game, Cooper expects that the cost will drop and the FDA and other regulators will loosen some of their requirements for clean facilities. Infection with pathogens from pig organs doesn't yet seem to be a problem, although Bennett and any future recipients will need to be monitored.

Nature **601**, 305-306 (2022)

doi: <https://doi.org/10.1038/d41586-022-00111-9>

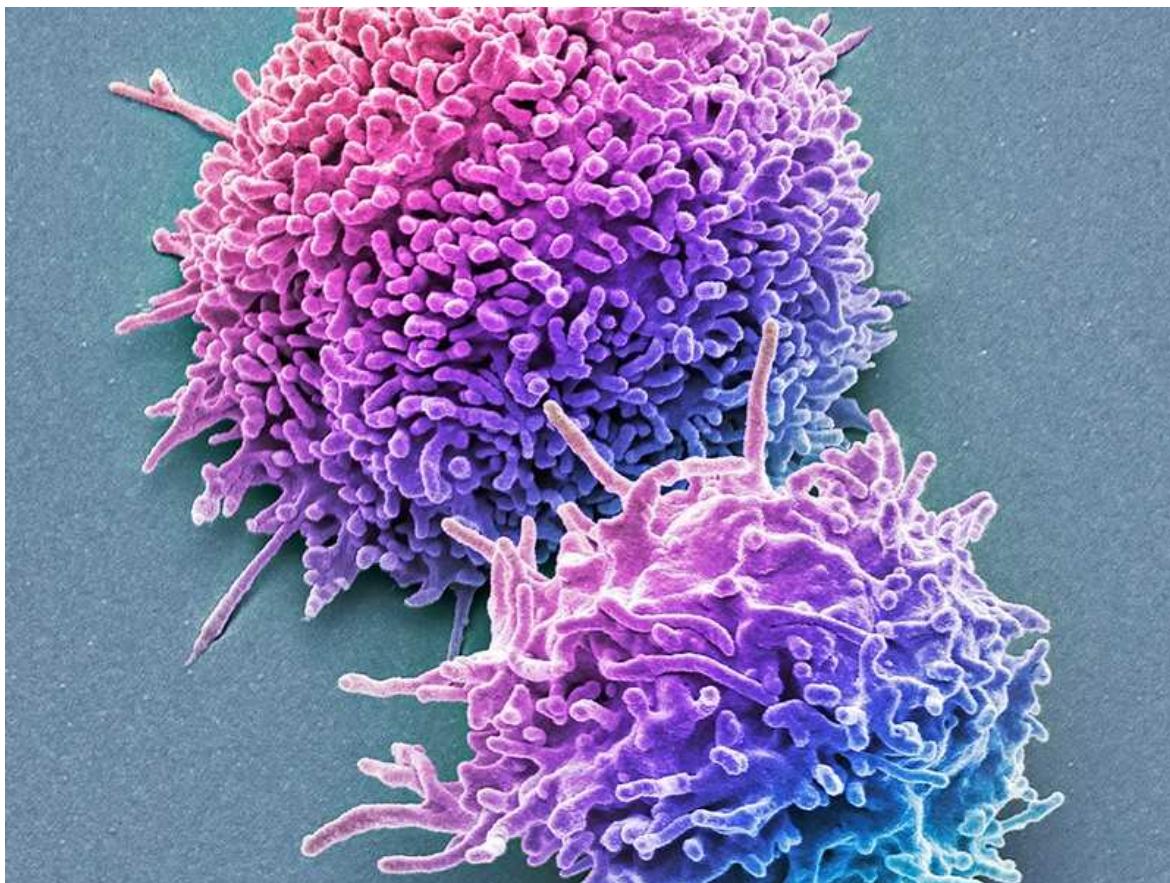
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00111-9>

- NEWS
- 11 January 2022

‘Killer’ immune cells still recognize Omicron variant

Amid concerns over lost antibody defences, some researchers argue that more attention should be paid to T cells.

- [Heidi Ledford](#)



T cells (seen in this scanning electron microscope image) from people who have been vaccinated or recovered from COVID-19 can recognize and target

the Omicron variant. Credit: Steve Gschmeissner/SPL

When immunologists Wendy Burgers and Catherine Riou heard about the Omicron coronavirus variant last November, they knew they would need to find the answers to some important questions. [Omicron's genome is loaded with mutations](#) — more than 30 in the region that codes for the spike protein, used in COVID-19 vaccines — meaning that the efficacy of antibodies raised against previous variants could be compromised.

Burgers and Riou, who both work at the University of Cape Town in South Africa, had studied previous variants and found that, although the emerging coronaviruses weakened people's antibody defences, another arm of the immune system — mediated by specialized cells called T cells — could still recognize the pathogens. But Omicron was more heavily mutated than any variant they'd studied before. How might this affect the immunity, hard won through vaccinations and previous infections, that populations were counting on to lessen the blow of COVID-19 surges? "With two to three times the amount of mutations, that made us think, 'Hmmm, we need to answer this question pretty quickly,'" says Burgers.

Since then, answers are beginning to come from a handful of laboratories around the world, and all converge on the same message. "The picture that's emerging is that [new] variants remain highly susceptible to T-cell responses," says Dan Barouch, director of the Center for Virology and Vaccine Research at Harvard Medical School in Boston, Massachusetts. "That includes Omicron."

Lasting immunity

When it comes to coronavirus immunity, antibodies have stolen the limelight. Researchers are monitoring people's levels of antibodies — particularly 'neutralizing antibodies' that directly prevent the virus from replicating — with bated breath. A drop in neutralizing-antibody levels correlates with [an increased risk of symptomatic infection](#). Antibodies are also easier to study than T cells, making it easier to analyse them in large, international vaccine trials.

But the rise of coronavirus variants has shown how fragile antibody-based immunity can be in the face of a changing virus. Neutralizing antibodies bind to a handful of regions on the SARS-CoV-2 spike protein, used as a template for many COVID-19 vaccines. Mutate those sites, and antibody protection fades.

T cells, however, are more resilient. These cells perform a variety of immune functions, including acting as ‘killer’ cells that destroy virus-infected cells. By killing infected cells, T cells can limit the spread of infection — and potentially reduce the chance of serious illness.

T-cell levels do not tend to fade as quickly as antibodies after an infection or vaccination. And because T cells can recognize many more sites along the spike protein than can antibodies, they are better able to recognize mutated variants. “What sounds like a lot of mutations doesn’t dent the T-cell response,” says Burgers.

So far, computer and lab analyses suggest that this is the case for Omicron. Several research groups have cross-referenced the mutations in Omicron with sites in the SARS-CoV-2 genome that are known targets of T cells. They have found that the majority of sites that T cells recognize are present in Omicron¹.

Other studies have analysed T cells taken from people who have either received a COVID-19 vaccine or been infected with a previous variant, and found that these T cells can respond to Omicron^{2–4}. “The T-cell responses remain quite intact, that’s good news,” says Corine Geurts van Kessel, a clinical virologist at the Erasmus Medical Centre in Rotterdam, the Netherlands. “The next step will be: what will it do in real life?”

Antibody focus

T-cell responses have been correlated with increased protection against severe COVID-19 in animal models and clinical studies in people. And Barouch suspects that T cells are responsible for the effectiveness of vaccines made by Pfizer–BioNTech⁵ and Janssen⁶ at preventing hospitalization due to an Omicron infection. “Neither of these vaccines

raised high levels of Omicron-neutralizing antibodies,” he says. “The efficacy data that we’re seeing coming out of South Africa, in my view, are likely due to T cells.”

That sometimes makes researchers’ heavy focus on antibody levels frustrating, says Harlan Robins, chief scientific officer and co-founder of Adaptive Biotechnologies, a company based in Seattle, Washington, that specializes in developing methods for studying T cells.

Last month, Pfizer and BioNTech announced that their COVID-19 vaccine failed to rouse a sufficient antibody response in children between two and five years old. As a result, the vaccine hasn’t been authorized in the United States for children under five. “They didn’t even look at T-cell response,” points out Robins.

And the large, initial vaccine trials in adults didn’t collect enough of the necessary samples to analyse whether T-cell responses can be correlated with vaccine efficacy. “It wasn’t scalable,” says Robins. “You can’t conduct a worldwide vaccine study and expect every group to viably preserve cells.” New, easier assays for studying T cells might make this more feasible in the future, he adds.

T cells might also get more attention as more variants emerge — and if the world begins to shift its focus from the number of infections to the severity of disease, says Geurts van Kessel. “Are you interested in infectivity? Then antibodies might be the more important measurement to do,” she says. “Whereas if you are interested in severe disease — which may be the case for the disease we are looking into now — T cells become much more important.”

Nature **601**, 307 (2022)

doi: <https://doi.org/10.1038/d41586-022-00063-0>

References

1. 1.

May, D. H. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.20.21267877> (2021).

2. 2.

Keeton, R. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.26.21268380> (2021).

3. 3.

Geurts van Kessel, C. H. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.27.21268416> (2021).

4. 4.

Liu, J. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2022.01.02.22268634> (2022).

5. 5.

Collie, S., Champion, J., Moultrie, H., Bekker, L.-G. & Gray, G. N.
Engl. J. Med. <https://doi.org/10.1056/NEJMc2119270> (2021).

6. 6.

Gray, G. E. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.28.21268436> (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00063-0>

- NEWS
- 12 January 2022

COVID vaccines safely protect pregnant people: the data are in

Despite evidence that pregnant people are at high risk of serious disease, many remain unvaccinated.

- [Shannon Hall](#)



A pregnant woman in Bogotá, Colombia, receives a COVID-19 jab in July.Credit: Raul Arboleda/AFP/Getty

COVID-19 can strike hard and fast — especially when you are pregnant. Alison Cahill, a specialist in maternal-fetal medicine at the Dell Medical

School in Austin, Texas, vividly remembers a patient from the first wave of the pandemic who was 26 weeks pregnant and woke up one morning with a cough. Her condition declined so rapidly that she was admitted to hospital that evening. Within six hours, she had been transferred to the intensive-care unit (ICU), where she was sedated so that she could be placed on a heart-lung bypass machine. Owing to safety precautions, her husband had to communicate with the medical team from the parking area.

“When she woke up and started to not feel very well, I don’t think it was within her wildest dreams that by the next morning she would be sedated and by herself in an ICU,” Cahill says. The woman spent a few weeks in the unit before she was finally able to go home.

Health-care professionals were still sharing similar gut-wrenching stories when the Delta variant of the coronavirus SARS-CoV-2 hit the United States. Cahill’s hospital, for example, was flooded with so many people who rapidly became ill with COVID-19 that the staff had to convert other floors into extra ICUs. But there was one stark difference: COVID-19 vaccines were now readily available. And all of Cahill’s critically ill pregnant patients had refused one.

When the vaccines were first administered to the public in late 2020, [little was known](#) about their effect on pregnant people, who hadn’t been represented in the original clinical trials that tested the shots. Although that is standard practice, it left pregnant people grappling with whether getting a jab was the best decision for themselves and their babies.

But the data are now solid. They show that the risks of COVID-19 during pregnancy — including maternal death, stillbirth and premature delivery — far outweigh the risks of being vaccinated. And yet, according to data from the US Centers for Disease Control and Prevention (CDC), only around 40% of pregnant people in the United States had been vaccinated against COVID-19 by 1 January 2022 (see ‘Slow uptake’). The numbers are similarly low across the globe — causing sick pregnant people to turn up at hospitals, and health-care workers to struggle to find solutions.

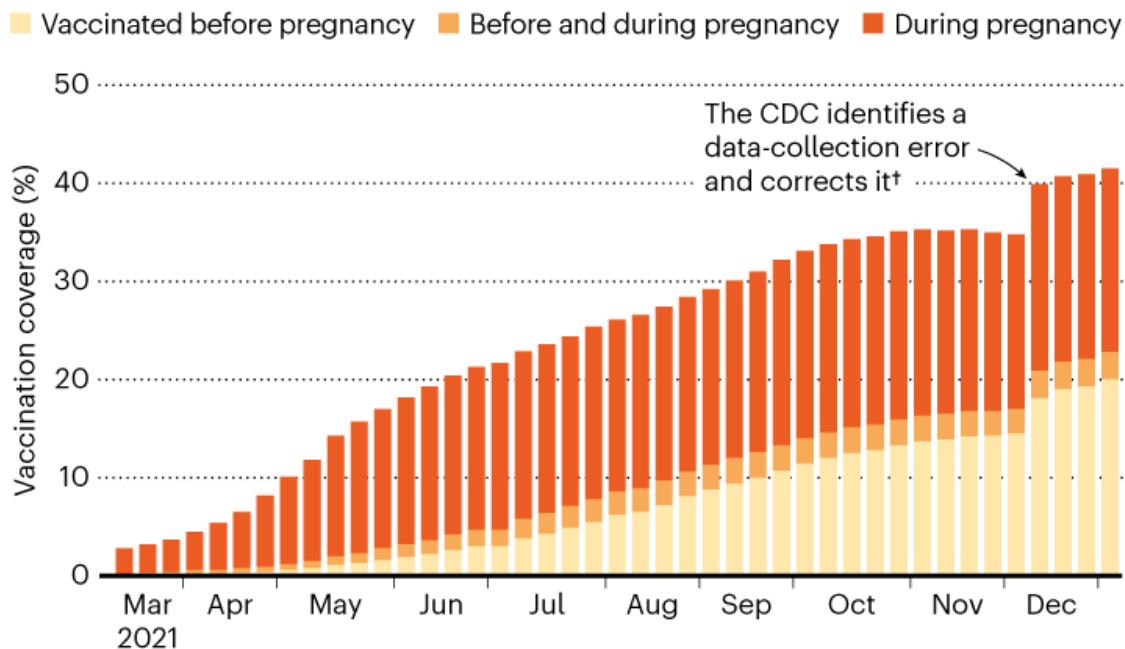
‘The perfect storm’

Whether COVID-19 poses a substantial risk during pregnancy was never in doubt. A pregnant person's body will suppress certain parts of the immune system to tolerate a fetus. Their blood volume also expands dramatically, straining the cardiovascular system. And the growing uterus pushes up on the diaphragm, reducing lung capacity.

"Then you take a viral illness that is potentially life threatening — one that affects the lungs and the cardiovascular system — and you have the perfect storm," says Andrea Edlow, a maternal–fetal medicine specialist at Harvard Medical School and Massachusetts General Hospital in Boston.

SLOW UPTAKE

By the end of 2021, a little more than 40% of pregnant people in the United States had been fully vaccinated* against COVID-19, compared with about 63% of the general US population, according to the US Centers for Disease Control and Prevention (CDC).



*Defined as having received two doses of an mRNA COVID-19 vaccine, or one dose of the Johnson & Johnson adenovirus vaccine.
†The CDC realized on 15 December 2021 that pregnant people who had received a COVID-19 booster hadn't been counted.

©nature

Source: Vaccine Safety Datalink/CDC

But the precise danger remained an open question until scientists could quantify it. In August 2021, a study¹ reviewed data from close to 870,000 women who gave birth at nearly 500 US medical centres between 1 March 2020 and 28 February 2021. The women who were diagnosed with COVID-19 were 15 times more likely to die and 14 times more likely to be intubated to help them breathe than were women without a COVID-19 diagnosis. They were also up to 22 times more likely to give birth prematurely.

The known risk, along with data demonstrating the vaccines' general safety, helped to nudge the CDC to change its tune on 11 August — recommending that pregnant people receive a vaccine. (Previously, the agency had simply advised that people at high risk of serious disease make the decision with the help of a physician.) The World Health Organization had recommended vaccines for the pregnant people at highest risk in June.

The change in guidelines came at a time when the Delta variant had changed the game: it caused an uptick in cases everywhere, but also an increase in the severity of illness among pregnant people. In October, for example, nearly 20% of the people most critically ill with COVID-19 in hospitals in England were unvaccinated pregnant women. Meanwhile, a CDC report² in late November found that the risk of stillbirth was 2.7 times higher among women who had a COVID-19 diagnosis at delivery during the Delta period than among those who had a diagnosis at delivery while previous strains were prevalent. And a second CDC report³ found that under the same conditions, the risk of maternal mortality was increased fivefold.

“If anything, COVID-19 is only becoming more deadly for the mother and the fetus,” Edlow says. (However, it’s not yet clear whether the Omicron SARS-CoV-2 variant, which emerged in November, has a similar impact.) By the end of 2021, more than 25,000 pregnant women in the United States had been admitted to hospital with the disease, and more than 250 had died.

A lifesaving shot

The increased risk encouraged more pregnant people (albeit a small percentage of the total population) to get inoculated when vaccines were

available to them — allowing monitoring systems around the globe to catch information about vaccination during pregnancy.

One study published in June 2021⁴ looked at adverse outcomes — pregnancy loss, including miscarriage and stillbirth, and neonatal outcomes such as premature birth, congenital disabilities and death — among 827 people in the United States who gave birth after receiving a COVID-19 vaccine. It found that they reported such events at a similar rate to people who gave birth before the pandemic (and had therefore not received a vaccine). And a study posted as a preprint in August⁵ looked at nearly 2,500 people who had received a COVID-19 vaccine either before conception or during the first 20 weeks of pregnancy, and found no increased risk of miscarriage.



A health-care worker tests a pregnant woman for COVID-19 in a suburb of Mumbai, India, in April. Credit: Satish Bate/Hindustan Times/Getty

Data have continued to pile up — showing that the vaccine is not only safe during pregnancy, but also effective. A study published in October⁶, for example, found that after the second dose of a vaccine, pregnant people mount a potent defence against the virus. And two studies^{7,8} published since

then found that vaccinated pregnant women were less likely to develop COVID-19 before delivery than were unvaccinated pregnant women. (Several formal clinical trials began last year to assess these questions.)

A bonus is that the vaccines might even protect the baby. Several studies published in the middle of last year^{9,10} suggested that antibodies generated in response to a COVID-19 vaccine are transferred through the placenta. That could confer immunity to the newborn — who is particularly vulnerable during the first few months of life and won't be eligible for a vaccine for some time. Although research on this subject is still scarce, Edlow's team reported¹¹ in a November preprint that antibodies persisted at 6 months of age in 60% of infants whose mothers were immunized against COVID-19 during pregnancy.

"I can't think of a better way to protect these babies once they're out in the world and potentially introduced to the COVID virus," says Jennifer Jolley, a maternal-fetal medicine specialist at the University of California, Irvine Medical Center.

Moving the needle

And yet inoculation rates among pregnant people are drastically low around the world. Not only are they below 50% in the United States, but an October meta-analysis found¹² that only 19% of pregnant people in Africa intend to receive a COVID-19 vaccine. Oceania has the highest intended uptake among pregnant people of any global region, with only 48%.

"It's getting increasingly hard to move the needle on this and increase those vaccination rates," says Sascha Ellington, an epidemiologist who leads the emergency-preparedness response team in the division of reproductive health at the CDC in Atlanta, Georgia.

Even before COVID-19, vaccine hesitancy during pregnancy was a known challenge. Before the H1N1 influenza pandemic that began in 2009, seasonal uptake of influenza vaccines among pregnant people in the United States hovered around 27%. But targeted strategies, such as an obstetrician recommending a shot or administering one during a prenatal appointment,

helped that to rise to 61% during the 2019–20 flu season — a figure comparable with that in the general adult population.

Many experts are hopeful that similar efforts will work for the COVID-19 vaccine. But the pandemic presents an entirely new challenge. That's in part due to the increasing power of social media to propagate misinformation, and the fact that differences in vaccine uptake have fallen along political lines. "It's a steeper uphill climb than ever before," Edlow says. But it's an important one. "It really is a matter of life or death."

Nature **601**, 308–309 (2022)

doi: <https://doi.org/10.1038/d41586-022-00031-8>

Editor's note: *Nature* recognizes that transgender men and non-binary people can become pregnant. We occasionally use 'women' in this story when discussing studies that used the same language.

References

1. 1.
Chinn, J. *et al.* *JAMA Netw. Open* **4**, e2120456 (2021).
2. 2.
DeSisto, C. L. *et al.* *Morb. Mortal. Wkly Rep.* **70**, 1640–1645 (2021).
3. 3.
Kasehagen, L. *et al.* *Morb. Mortal. Wkly Rep.* **70**, 1646–1648 (2021).
4. 4.
Shimabukuro, T. T. *et al.* *N. Engl. J. Med.* **384**, 2273–2282 (2021).
5. 5.

Zauche, L. H. *et al.* Preprint at Research Square
<https://doi.org/10.21203/rs.3.rs-798175/v1> (2021).

6. 6.

Atyeo, C. *et al.* *Sci. Transl. Med.* **13**, eabi8631 (2021).

7. 7.

Theiler, R. N. *et al.* *Am. J. Obstet. Gynecol. MFM* **3**, 100467 (2021).

8. 8.

Morgan, J. A. *et al.* *Obstet. Gynecol.* **139**, 107–109 (2022).

9. 9.

Gray, K. J. *et al.* *Am. J. Obstet. Gynecol.* **225**, P303.E1-303.E17 (2021).

10. 10.

Beharier, O. *et al.* *J. Clin. Invest.* **131**, e150319 (2021).

11. 11.

Shook, L. L. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.11.17.21266415> (2021).

12. 12.

Shamshirsaz, A. A. *et al.* *Am. J. Perinatol.* <https://doi.org/10.1055/a-1674-6120> (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00031-8>

- NEWS
- 11 January 2022
- Update [11 January 2022](#)

How to protect US science from political meddling after Trump

In a fresh report, federal researchers recommend ways to strengthen scientific integrity and preserve public trust in government.

- [Alexandra Witze](#)



During a 2019 media briefing, then-president Donald Trump contradicted a federal weather office by saying that Hurricane Dorian would affect Alabama. Credit: Tom Brenner/Bloomberg/Getty

US federal agencies need to strengthen the policies that protect science used in government decision-making. They should also create a scientific-integrity council spanning many agencies, to help address political meddling by government officials. Those are some of the top conclusions of a long-awaited report from a task force convened by US President Joe Biden's administration, in the wake of four years of science — and scientists — being undermined under former president Donald Trump.

The White House Office of Science and Technology Policy (OSTP) [released the report](#), which tackles the thorny topic of scientific integrity in the federal government, on 11 January. High-profile examples of scientific integrity being violated include a 2019 incident in which Trump said that a hurricane was threatening parts of Alabama, even though a weather-forecast office said it wasn't; the acting head of the National Oceanic and Atmospheric Administration was then involved in a statement criticizing government scientists for contradicting Trump. "Political intrusion into the conduct, management, communication, and use (or misuse) of science has a severe impact on public trust in federal science," says the report.

The report calls on agencies to increase accountability for those found to have violated scientific-integrity principles. It also recommends updating policies to incorporate emerging technologies such as artificial intelligence and machine learning.

"It really is amazing to see the Biden administration take this issue so seriously, especially after what we witnessed" under Trump, says Jacob Carter, a senior scientist with the Union of Concerned Scientists, who is based in Washington DC and studies scientific integrity. But several aspects need more work, he says: for example, agencies need to flesh out what accountability would really look like.

Restoring trust

The task force includes around 50 experts from 29 government agencies, spearheaded by Alondra Nelson, the OSTP's deputy director for science and society, and Jane Lubchenco, the office's deputy director for climate and environment. It was created after Biden issued a memorandum in January

2021, one week into his presidency, calling for “restoring trust in government through scientific integrity and evidence-based policymaking”. The task force began its work in May 2021 and, [following public meetings and listening sessions](#), has now issued its full report.

It is a “superb job”, says John Holdren, who headed the OSTP between 2009 and 2017 as former president Barack Obama’s science adviser. “It represents an entirely new standard on this critical issue.”

The report’s authors say this is only the first step: future work will involve putting the recommendations into practice across government in the coming months. “It is a process,” Nelson told a meeting of the American Geophysical Union (AGU) last month. “We will always be working on scientific integrity.”

The question now is whether federal agencies can strengthen their scientific-integrity policies to be truly effective against challenges such as those that arose during Trump’s presidency. “By elevating the importance of this issue ... it is our hope that we can minimize the likelihood of future violations,” Lubchenco says.

Many agencies already had such policies when Trump took office in 2017. They simply were not strong enough to withstand the suppression of science that occurred. One major problem was that top officials were sometimes involved in integrity breaches — and agencies struggled with how to respond to such events. For example, in 2017, Scott Pruitt, then administrator of the Environmental Protection Agency, stated that carbon dioxide was not a primary contributor to global warming. Agency officials investigated whether this violated the agency’s scientific-integrity policy, but ultimately cleared Pruitt. “Violations involving high-level officials are the most problematic and difficult to address,” says the OSTP task force in its report.

The report recommends establishing a formal Scientific Integrity Council, with members drawn from many agencies. Such a council might deal better with possible violations involving top officials, because it would provide a mechanism for investigation outside the agency involved. Other suggested changes include standardizing policies across agencies and expanding

scientific-integrity training to include contractors and scientists who receive agency grants — both of which could help to ingrain awareness of scientific integrity more widely, Carter says.

Perhaps most importantly, those found to have violated scientific-integrity policies need to be held to account, the report says — with consequences comparable to those for violating government ethics rules. It does not, however, lay out what such consequences might look like.

“This is a situation where the devil is truly in the details, and the lack of specificity is frustrating,” says Lauren Kurtz, executive director of the Climate Science Legal Defense Fund in New York City.

Lubchenco says that more details will be forthcoming about what such consequences might involve. “There should not just be a minor slap on the hand,” she says.

Overriding decisions

The report recommends broadening scientific-integrity policies beyond science agencies such as the National Institutes of Health and the National Science Foundation. For instance, the departments of justice and transportation use science to inform decisions on issues such as incarceration and urban planning. “We need to engage everyone who communicates about science,” Nelson told the AGU meeting.

The task force’s work builds on the OSTP’s efforts under the Obama administration. In 2010, Holdren [issued a short but influential memo](#) that laid out basic guidelines for scientific integrity in the federal government. However, those guidelines didn’t stand up to the Trump administration. One database, the Silencing Science Tracker — run by the Sabin Center for Climate Change Law at Columbia University in New York City and the Climate Science Legal Defense Fund — found that the federal government took 339 anti-science actions during the Trump administration. Only nine of those have been reviewed or reversed by the Biden administration, Kurtz says.

Each president can override many of the decisions made by the previous incumbent. So, for changes to scientific-integrity policies to be permanent, they will have to be enacted by legislators in Congress, Carter says. “It will be incredibly important to codify a lot of this in legislation,” he says.

Congress is currently considering bills that would require agencies to adopt uniform policies on scientific integrity.

Nature **601**, 310-311 (2022)

doi: <https://doi.org/10.1038/d41586-022-00059-w>

Updates & Corrections

- **Update 11 January 2022:** This story was updated to include a comment from former OSTP head John Holdren.

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00059-w>

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

- NEWS
- 13 January 2022
- Correction [13 January 2022](#)

Omicron thwarts some of the world's most-used COVID vaccines

Inactivated-virus vaccines elicit few, if any, infection-blocking antibodies — but might still protect against severe disease.

- [Elie Dolgin](#) ⁰



A man in Javanese traditional dress receives a dose of the Sinovac COVID-19 vaccine at a temple in Yogyakarta, Indonesia. Credit: Ulet Ifansasti/Getty

The world's most widely used COVID-19 vaccines provide little to no protection against infection with the rapidly spreading [Omicron variant](#), laboratory evidence suggests.

Inactivated-virus vaccines contain SARS-CoV-2 particles that have been chemically treated to make it impossible for them to cause an infection. Stable and relatively easy to manufacture, such vaccines have been distributed widely as part of [China's global vaccine diplomacy](#), helping them to become the jab of choice in many countries. But a multitude of experiments show that they are consistently hobbled by Omicron.

Many people who receive two jabs of an inactivated vaccine fail to produce immune molecules that can counter Omicron transmission. And even after a third dose of an inactivated vaccine, an individual's levels of 'neutralizing' antibodies, which provide a potent safeguard against viral infection of cells, tend to remain low. A third shot of another type of vaccine, such as those based on [messenger RNA](#) or [purified proteins](#), seems to offer better protection against Omicron.

The findings are prompting many scientists and public-health researchers to re-evaluate the role of inactivated vaccines in the global fight against COVID-19.

"At this stage, we have to evolve our ideas and adjust our vaccination strategies," says Qiang Pan-Hammarström, a clinical immunologist at the Karolinska Institute in Stockholm.

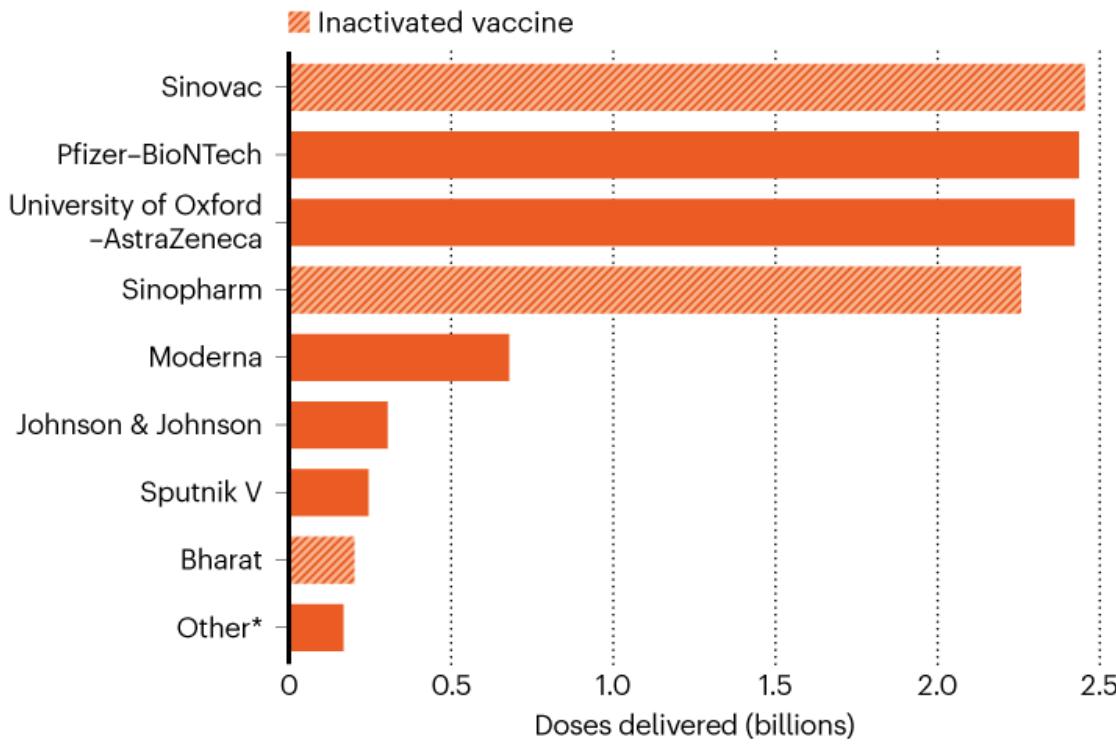
Billions served

Inactivated vaccines were [instrumental in the campaign for worldwide vaccine coverage](#) last year. They include those made by China's Sinovac and Sinopharm, which together account for nearly 5 billion of the more than 11 billion COVID-19 vaccine doses delivered globally so far, according to numbers compiled by data-tracking firm Airfinity in London (see 'Many shields against COVID-19'). More than 200 million doses of other

inactivated shots such as India's Covaxin, [Iran's COVI](#)[Iran Barekat](#) and Kazakhstan's QazVac have also been delivered.

MANY SHIELDS AGAINST COVID-19

More than 11 billion doses of COVID-19 vaccine have been delivered around the world. Roughly half are the inactivated vaccines made by Sinopharm, Sinovac and Bharat Biotech.



*Includes inactivated vaccines and other types.

©nature

Source: Data from Airfinity

Such products remain crucial for preventing hospitalization and death from COVID-19. And they can still serve a valuable immune-priming function for as-yet unvaccinated individuals.

But an early sign that inactivated vaccines might not hold up to Omicron came in December, when researchers in Hong Kong analysed blood from 25 recipients of the two-dose [CoronaVac vaccine, made by the Beijing-based company Sinovac](#). Not a single person had detectable neutralizing antibodies against the new variant — raising the possibility that all the participants were highly vulnerable to Omicron infection¹.

Sinovac has disputed this finding, pointing to internal data showing that 7 out of 20 people who had received the company's vaccine had tested positive for antibodies capable of neutralizing Omicron. Other studies involving people immunized with [Covaxin²](#), which is made by Bharat Biotech in Hyderabad, India, and BBIBP-CorV³, produced by state-owned Chinese company Sinopharm, in Beijing, have also concluded that inactivated vaccines retain some potency against Omicron — although, as researchers at the Translational Health Science and Technology Institute in Faridabad, India, put it in their study², the immune responses remain “sub-optimal”. The work on Covaxin has not yet been peer reviewed.

Immunity top-up

A third dose of inactivated vaccine helps to restore neutralization activity for many individuals. A 292-person study by researchers at the Shanghai Jiao Tong University School of Medicine in China, for example, identified neutralizing antibodies against Omicron in just 8 people tested 8–9 months after an initial course of BBIBP-CorV. After another shot of the same vaccine, that number rose to 228⁴. This work has not yet been peer reviewed.

Levels of neutralizing antibodies in each person's blood remained low. But as molecular virologist Rafael Medina at the Pontifical Catholic University of Chile in Santiago points out: “There are other parts of the immune response that are also playing a role.” [T cells destroy infected cells; B cells remember past infections](#) and strengthen immune responses for the future; and binding antibodies contribute to viral control.

In a preprint published in December⁵, Medina and his co-authors — led by immunologist Galit Alter at the Ragon Institute of MGH, MIT and Harvard in Cambridge, Massachusetts — showed that people immunized with CoronaVac maintain non-neutralizing antibodies that both bind Omicron and assist immune cells in gobbling up infected cells.

On the defensive

Those kinds of result show that recipients of inactivated vaccines, although not necessarily protected against infection by Omicron, should still be shielded from the worst ravages of COVID-19 triggered by the variant, says Murat Akova, an infectious-disease specialist at Hacettepe University School of Medicine in Ankara.

All the same, an extra dose of vaccine could offer some much-needed immune insurance. Experiments conducted by Pan-Hammarström and her colleagues found that, after two doses of inactivated vaccine, an mRNA top-up hoists levels of binding antibodies, memory B cells and T cells⁶. And studies of samples from China^{3,7} and the United Arab Emirates⁸ have shown that a protein-based booster triggers higher numbers of neutralizing antibodies than does a third shot of an inactivated vaccine. Many of these results have not yet been peer reviewed.

Double boost?

But [a single booster with a different type of vaccine](#) might not be enough to subdue Omicron, warns Akiko Iwasaki, a viral immunologist at Yale School of Medicine in New Haven, Connecticut.

Iwasaki and her co-authors studied blood samples from 101 individuals who received two doses of CoronaVac followed by an mRNA booster. Before the boost, the samples showed no detectable Omicron neutralization. Afterwards, 80% of analysed samples showed some Omicron-blocking activity⁹. But the quantities of antibodies that had Omicron-neutralizing potential were not much greater in this group than in a separate population that had received two doses of mRNA vaccine and no booster. The work has not yet been peer reviewed.

Before the Omicron variant emerged, Iwasaki had been advocating single [mRNA boosters](#) for recipients of inactivated vaccines. “We were really celebrating how wonderful this strategy is,” she says, “and then — boom! — Omicron hit.” Now, she thinks these people probably need two extra jabs.

“The bar keeps being raised by the variants,” Iwasaki says. “We’re playing catch up all the time.”

Nature **601**, 311 (2022)

doi: <https://doi.org/10.1038/d41586-022-00079-6>

Updates & Corrections

- **Correction 13 January 2022:** The original version of the graphic showed an incorrect number of vaccine doses delivered in the ‘Other’ category. This has now been corrected.

References

1. 1.

Lu, L. *et al.* *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciab1041> (2021).

2. 2.

Medigeshi, G. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2022.01.04.22268747> (2022).

3. 3.

Ai, J. *et al.* *Emerg. Microbe. Infect.*
<https://doi.org/10.1080/22221751.2021.2022440> (2021).

4. 4.

Yu, X. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.17.21267961> (2021).

5. 5.

Bartsch, Y. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.24.21268378> (2021).

6. 6.

Zuo, F. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2022.01.04.22268755> (2022).

7. 7.

Wang, X. *et al.* Preprint at bioRxiv
<https://doi.org/10.1101/2021.12.24.474138> (2021).

8. 8.

AlKaabi, N. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.29.21268499> (2022).

9. 9.

Pérez-Then, E. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.27.21268459> (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00079-6>

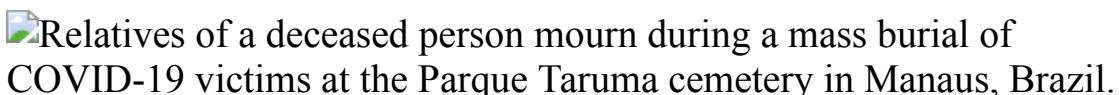
| [Section menu](#) | [Main menu](#) |

- NEWS FEATURE
- 18 January 2022

The pandemic's true death toll: millions more than official counts

Countries have reported some five million COVID-19 deaths in two years, but global excess deaths are estimated at double or even quadruple that figure.

- [David Adam](#) ⁰



A family stands at the grave of a relative who died from COVID-19 in Manaus, Brazil, in May 2020. Credit: Andre Coelho/Getty

Last year's Day of the Dead marked a grim milestone. On 1 November, the global death toll from the COVID-19 pandemic passed 5 million, official data suggested. It has now reached 5.5 million. But that figure is a significant underestimate. Records of excess mortality — [a metric that involves comparing all deaths recorded with those expected to occur](#) — show many more people than this have died in the pandemic.

Working out how many more is a complex research challenge. It is not as simple as just counting up each country's excess mortality figures. Some official data in this regard are flawed, scientists have found. And more than 100 countries do not collect reliable statistics on expected or actual deaths at all, or do not release them in a timely manner.

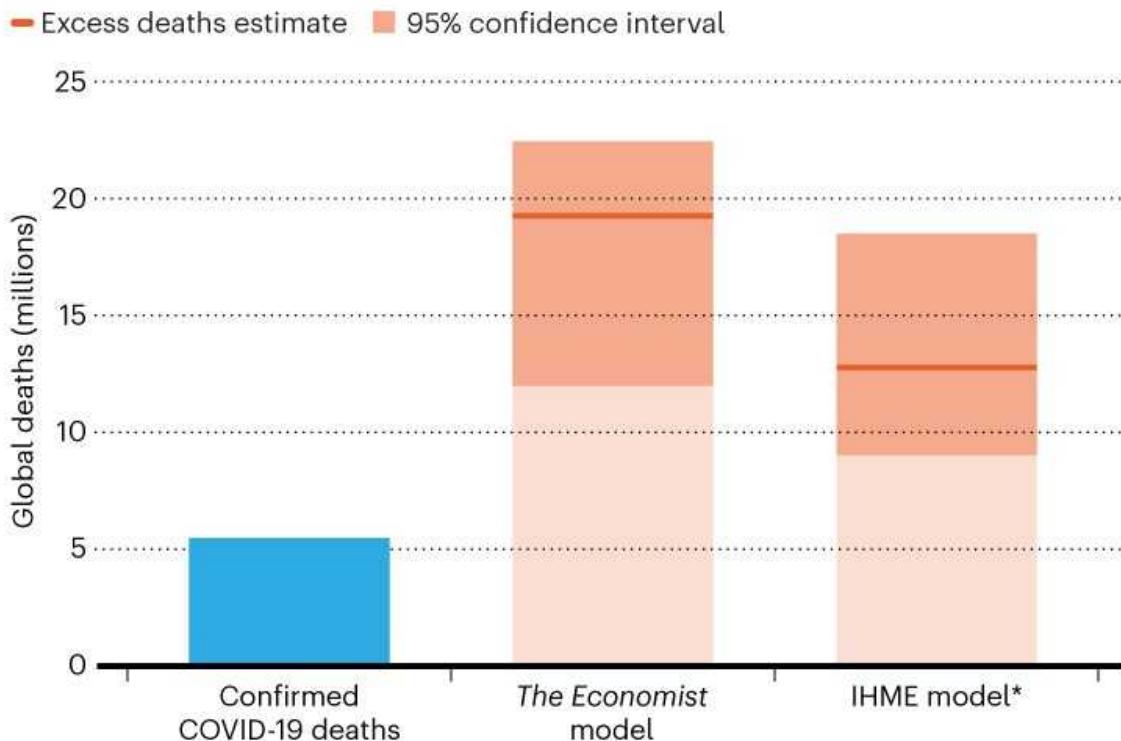
Demographers, data scientists and public-health experts are striving to narrow the uncertainties for a global estimate of pandemic deaths. These efforts, from both academics and journalists, use methods ranging from

satellite images of cemeteries to door-to-door surveys and machine-learning computer models that try to extrapolate global estimates from available data.

Among these models, the World Health Organization (WHO) is still working on its first global estimate, but the Institute for Health Metrics and Evaluation in Seattle, Washington, [offers daily updates](#) of its own modelled results, as well as projections of how quickly the global toll might rise. And one of the highest-profile attempts to model a global estimate has come from the news media. *The Economist* magazine in London has used a machine-learning approach to produce an estimate of 12 million to 22 million excess deaths — or between 2 and 4 times the pandemic's official toll so far (see go.nature.com/3qjtyge and 'Global toll').

GLOBAL TOLL

By January 2022, there had been 5.5 million official COVID-19 deaths worldwide in the pandemic. But models estimate that there have been between two and four times that number of excess deaths — that is, mortality above what was expected — since the start of 2020.



Sources: Our World in Data/*The Economist*/IHME

The uncertainty in this estimate is a discrepancy the size the population of Sweden. “The only fair thing to present at this point is a very wide range,” says Sondre Ulvund Solstad, a data scientist who leads *The Economist*’s modelling work. “But as more data come in, we are able to narrow it.”

The scramble to calculate a global death toll while the pandemic continues is an exercise that combines sophisticated statistical modelling with rapid-fire data gathering. Everyone involved knows any answer they provide will be provisional and imprecise. But they feel it is important to try. They want to acknowledge the true size and cost of the human tragedy of COVID-19 (see ‘Comparing pandemics’), and they hope to counter misleading claims prompted by official figures, such as China’s count of just under 5,000 COVID-19 deaths.

Comparing pandemics

According to some estimates of excess deaths, the COVID-19 pandemic is the largest since the 1918–20 H1N1 influenza pandemic when scaled to 2020 populations.

Metric	COVID-19	Flu 2009 (H1N1)	Flu 1968 (H3N2)	Flu 1957–59 (H2N2)	Flu 1918–20 (H1N1)
Per-capita excess mortality rate (estimate)	0.15–0.28%	0.005%	0.03%	0.04%	1%
Global excess deaths (estimate) adjusted to 2020 population	12 million–22 million	0.4 million	2.2 million	3.1 million	75 million

Metric	COVID-19	Flu 2009 (H1N1)	Flu 1968 (H3N2)	Flu 1957–59 (H2N2)	Flu 1918–20 (H1N1)
--------	----------	-----------------	-----------------	--------------------	--------------------

Mean age at death (years;
United States and Europe
only) 73–79 37 62 65 27

Sources: [Simonsen, L. & Viboud, C. eLife 10, e71974 \(2021\)](#); COVID-19 estimates: *The Economist's* model (to January 2022); age of death data: US CDC, UKHSA.

Flawed figures

Death and taxes are famously the only certainties in life, but countries account for each of them in vastly different ways. Even superficially similar places can have varying approaches to recording COVID-19 deaths. Early in the pandemic, countries such as the Netherlands counted only those individuals who died in hospital after testing positive for the coronavirus SARS-CoV-2. Neighbouring Belgium included deaths in the community and everyone who died after showing symptoms of the disease, even if they weren't diagnosed.

That is why researchers quickly turned to excess mortality as a proxy measure of the pandemic's toll. Excess-death figures are seemingly easy to calculate: compare deaths during the pandemic with the average recorded over the previous five years or so. But even in wealthy countries with comprehensive and sophisticated systems to report deaths, excess-mortality figures can be misleading. That's because the most obvious way to calculate them can fail to account for changes in population structure.

“We should be careful about this issue, because looking at the average raw data is really flawed,” says Giacomo De Nicola, a statistician at Ludwig Maximilian University of Munich, Germany.

When De Nicola and colleagues worked on a 2021 study to calculate excess mortality caused by the pandemic in Germany, they found that comparing deaths to average mortality in previous years consistently underestimated the number of expected deaths, and so overstated excess deaths¹. The reason was a rise in annual national mortality, contributed to by a surge in the number of people aged 80 and above — a generation too young to fight and die in the Second World War.

The difference for Germany is significant. Press-released raw data from the German statistical office last year reported 5% more deaths in 2020 compared with 2019. But after taking the age structure into account, De Nicola's group reduced this to just 1%. "Due to the lack of a generally accepted method for age-adjustment, I'm pretty certain this issue extends to many more countries," he says.

Some demographers agree. "It concerns me that some so-called excess-deaths estimates by national statistical offices just use an average of the past five years of deaths as the expected deaths. In ageing populations, this is unlikely to be the best estimate," says Tom Wilson, a demographer at the University of Melbourne, Australia. Responding to De Nicola's work, Felix zur Nieden, a demographer at Germany's statistical office, says he agrees that raw numbers should be adjusted to take age structure and other subtleties into account.

More-sophisticated analyses adjust the expected deaths baseline to account for such biases, for example by raising the number of expected deaths as a population ages. Probably the most comprehensive of these excess-mortality estimates come from Ariel Karlinsky, an economist at the Hebrew University of Jerusalem in Israel, and Dmitry Kobak, a data scientist at the University of Tübingen, Germany.

Since January 2021, Karlinsky and Kobak have produced a regularly updated database of all-cause mortality before and during the pandemic (2015–21) from as many sources and for as many places as possible² — currently some 116 countries and territories. Called [the World Mortality Dataset \(WMD\)](#), the bulk of the information comes from official death statistics collected and published by national offices and governments. The duo then works with these data to [estimate excess mortality](#), including trying

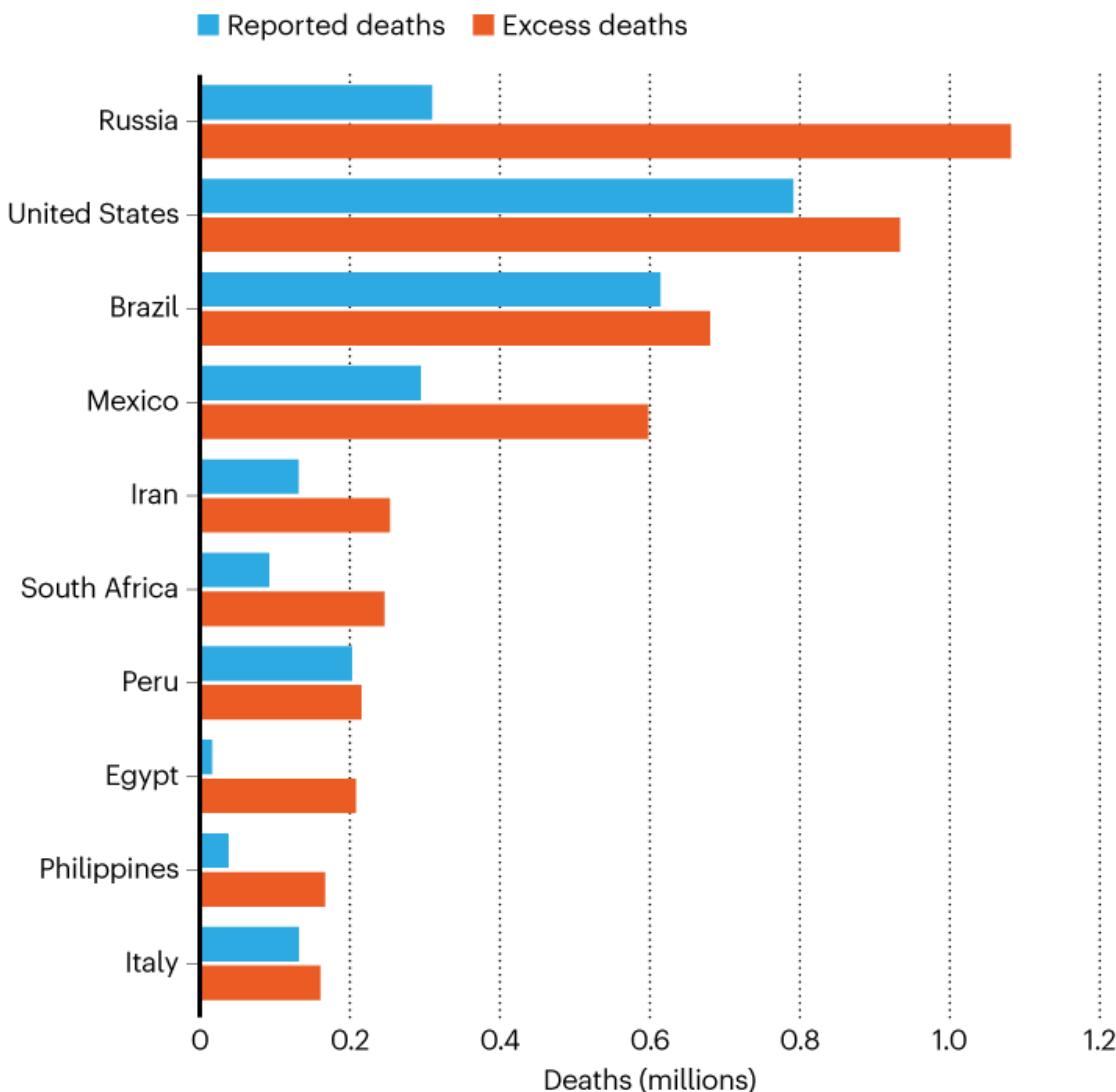
to take into account death tolls associated with armed conflict, natural disasters and heatwaves. For example, they assumed that 4,000 lives were lost in both Armenia and Azerbaijan during the 2020 Nagorno-Karabakh war.

Karlinsky, who previously worked on health economics, recognized that even the best epidemiological models were based on official reported COVID-19 numbers that, for many places, were clearly too low or missing entirely. “Many people had been throwing around their conjectures about excess mortality without basing it on data,” he says.

In many cases, Karlinsky and Kobak’s estimates of excess deaths diverge significantly from COVID-19 mortality statistics released by governments. Russia, for instance, reported more than 300,000 COVID-19 deaths by the end of 2021, but is likely to have exceeded 1 million excess deaths in that time (see ‘Excess deaths’).

EXCESS DEATHS

The 116 countries and territories in the World Mortality Dataset have reported 4.1 million COVID-19 deaths so far, but their combined excess mortality is around 1.6 times higher, at 6.5 million. Ten example countries are shown.



Data collated up to 31 Dec 2021 (Russia); 12 Dec 2021 (United States); 30 Nov 2021 (Brazil); 14 Nov 2021 (Mexico); 2 Jan 2022 (Iran); 9 Jan 2022 (South Africa and Peru); 30 Jun 2021 (Egypt); 30 Sept 2021 (Philippines); 31 Oct 2021 (Italy). The Economist's model gives similar figures because it draws on World Mortality Dataset numbers for excess deaths, but adds modelling to bring results to the present day.

©nature

Source: World Mortality Dataset

For countries covered by the WMD, official figures suggest that 4.1 million deaths since the start of the pandemic are down to COVID-19 — around

10% of all deaths during that time. But the duo's calculations suggest that, when excess mortality is taken into account, deaths related to COVID-19 are 1.6 times greater, at around 6.5 million deaths (or 16% of the total). In some countries, the relative impact of the virus is even higher. One-third of all deaths in Mexico can be attributed to the virus, Karlinsky and Kobak's data suggest.

Excess deaths include mortality that is not related to COVID-19, such as other infectious diseases, as well as indirectly related deaths, such as a person with cancer who died because their screening was cancelled owing to the pandemic's impact on health-care systems. Some countries, such as New Zealand, even had negative excess mortality, because they experienced few losses to COVID-19 and saw a drop in deaths from influenza. But Karlinsky argues that, overall, data show that estimating excess deaths is a reliable way to measure COVID-19 casualties.

Modelling global deaths

The WMD lacks excess-death estimates for more than 100 countries, including China, India and many in Africa. That's because those countries either do not collect death statistics or do not publish them speedily. But they also account for millions of COVID-19 deaths. A true pandemic global death toll cannot be counted without those data, but some researchers argue it is possible to model one.

Such an estimate has been produced for a pandemic before — for influenza. Starting in the Americas in March 2009, a type of H1N1 influenza A virus ravaged the world for more than a year. By the time the WHO declared that pandemic over in August 2010, the organization's 'official' toll, made up of laboratory-confirmed deaths, was less than 19,000.

A team of international public-health experts took a different approach. Starting with estimated influenza deaths in 20 countries, together covering more than one-third of the world's population, the researchers looked for factors that could explain why some of these countries fared better or worse than others. They found ten indicators, including population density, number of doctors and income. The relationship between these contributing factors

and deaths for a given country allowed them to model how many flu deaths they would expect in other countries, purely on the basis of a country's performance on these indicators³.

Their study suggested that between 123,000 and 203,000 people died in the pandemic in the last 9 months of 2009 — about 10 times the WHO count. In 2019, the same team repeated the exercise to model deaths from seasonal flu epidemics from 2002 to 2011, starting this time with data from 31 countries. They reported that an average of 389,000 respiratory deaths (uncertainty range 294,000 to 518,000) were associated with flu globally for each year modelled⁴.

The same method should work for COVID-19, says Cécile Viboud, an epidemiologist at the National Institutes of Health in Bethesda, Maryland, who worked on the 2019 influenza study. “We have much more data [for COVID-19] than we did with flu. So, in a way it is cleaner.” Unlike with flu, it should be much easier to attribute respiratory deaths to the COVID-19 pandemic, she says, because the circulation of almost every other respiratory pathogen was stopped owing to lockdowns and other measures.

“Statistically, it’s a much easier proposition,” Viboud says.

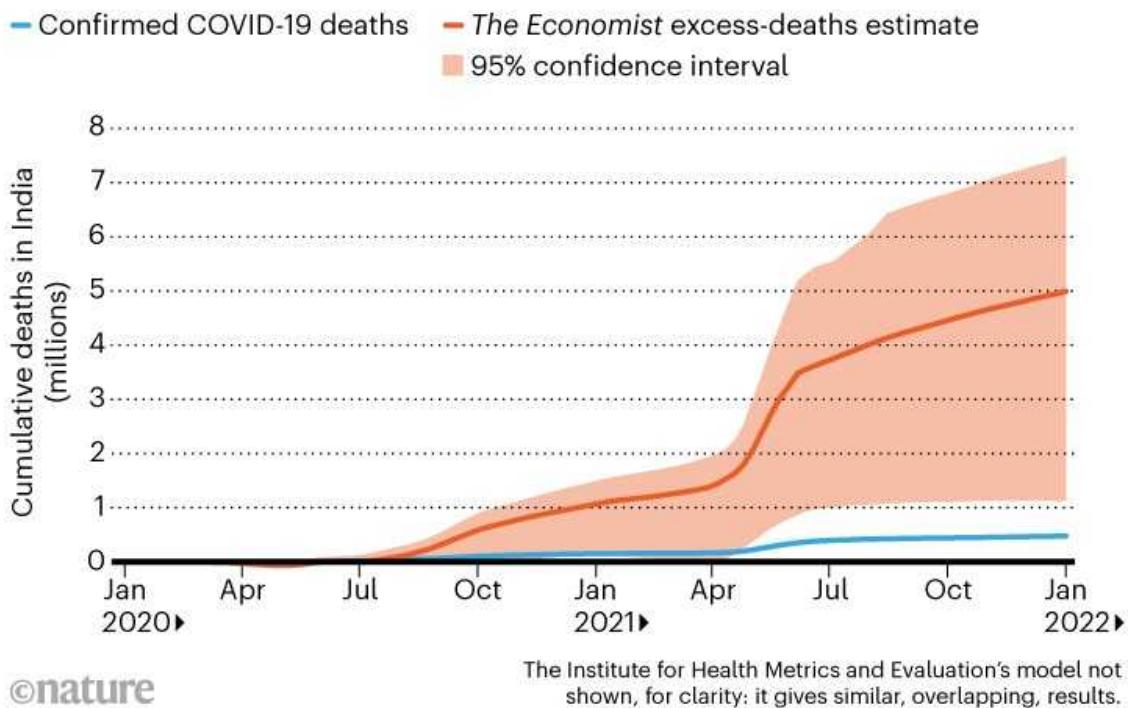
The model used by *The Economist* to track the COVID-19 pandemic uses machine learning to identify more than 100 national indicators that seem to correlate with excess deaths in more than 80 countries where data are available. These features include official deaths, the scale of COVID-19 testing and the results of antibody surveys, but also geographical latitude, the degree of Internet censorship and the number of years a country has been a democracy. It is possible to examine the importance of each indicator in the model, but this is far from straightforward — features can act in combination, and their relative importance might differ for countries that have different characteristics, says Solstad.

Plug numbers for these indicators for a country that doesn’t produce mortality data into the model, and algorithms estimate that country’s excess deaths. The model estimates some 5 million deaths in India, for example, 10 times higher than the country’s official COVID-19 toll of less than 500,000 deaths (see ‘Millions of missing deaths’). That estimate is sadly plausible — using sample surveys of households and sub-national mortality data,

academic groups have separately estimated that as many as 3 million to 5 million people might have died from COVID-19 in India^{5,6}. *The Economist*'s algorithm has a wide uncertainty interval of between 1 million and 7.5 million deaths for India.

MILLIONS OF MISSING DEATHS

India has reported fewer than 500,000 COVID-19 deaths — but both *The Economist*'s model and household surveys suggest that real pandemic toll is in the millions.



Sources: Our World in Data/*The Economist*

For China, the model estimates almost 750,000 deaths (well over 150 times higher than the country's reported 4,600), but with a wide uncertainty interval ranging from as low as 200,000 fewer deaths than expected, to as high as 1.9 million excess deaths.

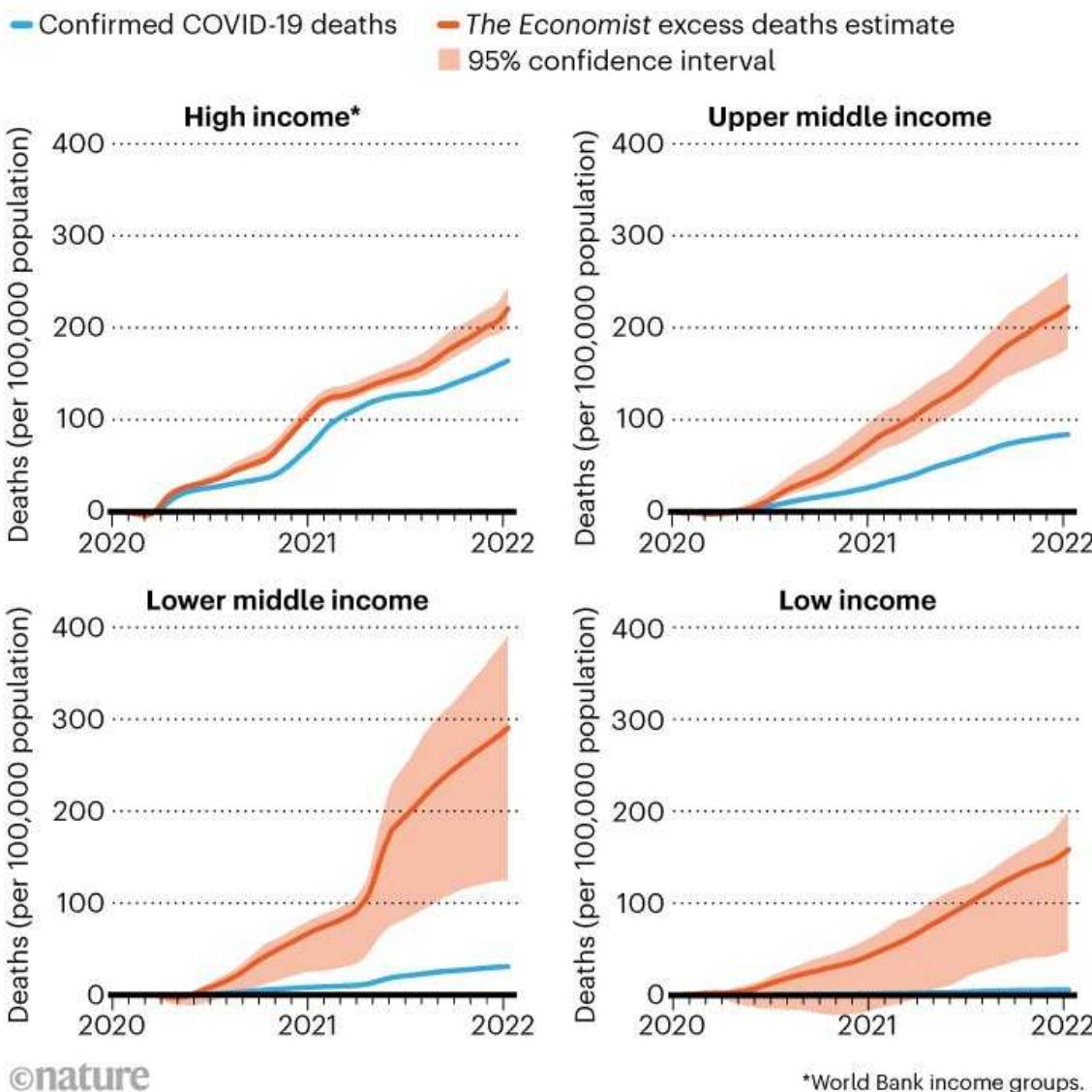
The Economist's model highlights how countries' official death counts often underestimate the true number — but that the extent of the underestimate varies. Excess deaths in the world's richest countries might be around one-

third above official counts, but those in the poorest countries could be more than 20 times higher, although these estimates are extremely uncertain.

Overall, the model suggests that lower middle-income countries (as described by World Bank groupings) have suffered at least as severely in per-capita deaths as rich countries — in contrast to the picture given by official figures (see ‘Rich and poor’). That’s despite the fact that these poorer countries have younger populations, adds Solstad.

RICH AND POOR

Official figures suggest that wealthy countries had the highest number of deaths per capita during the pandemic. But a model that estimates excess deaths suggests that is false: lower middle-income countries might have been hit hardest.



Source: *The Economist*

Bad practice?

Not everyone agrees with the approach. One vocal critic of the magazine's pandemic modelling is Gordon Shotwell, a data scientist in Halifax, Nova

Scotia, who published a blog post that called it irresponsible (see go.nature.com/3jpdkrs). “Models like this have the effect of putting a thin veneer of objectivity and science-y thinking over what’s basically an op-ed,” he wrote.

In September, for instance, the magazine used its model results to say that pandemic deaths in Kenya were between 19,000 and 110,000, versus an official figure of 4,746.

“Using any model to make an estimate about those places I think is just bad practice,” Shotwell told *Nature*. “You don’t learn anything by training a model on mostly rich countries with high life expectancy and applying it to poor countries with low life expectancy.”

Solstad, not surprisingly, sees it differently: “I think it is better to provide an uncertain number than to rely on a very certain number that is clearly false.”

Very low or zero ‘official’ numbers of COVID-19 deaths for countries where data are patchy or lacking present problems of their own, he says. They have fuelled nonsense theories that people in Africa have genetic resistance to the disease and don’t need international help or vaccines, for instance.

Some demographers see Shotwell’s point of view, saying that applying modelling to countries without their own deaths data is inherently difficult. “The process is intrinsically flawed. The data are a real mess and so any modelling effort is going to be very speculative,” says Jon Wakefield, a statistician at the University of Washington in Seattle, who leads a modelling project run by the WHO to estimate the pandemic’s excess death toll. “It’s very frustrating as the data are so limited. I’m not happy with the assumptions we’re being forced to make, but we’re doing the best we can.”

The project, which uses a more straightforward statistical model than *The Economist* to fill in the gaps, was scheduled to publish its first results in December, but they had not been released by mid-January as *Nature* went to press.

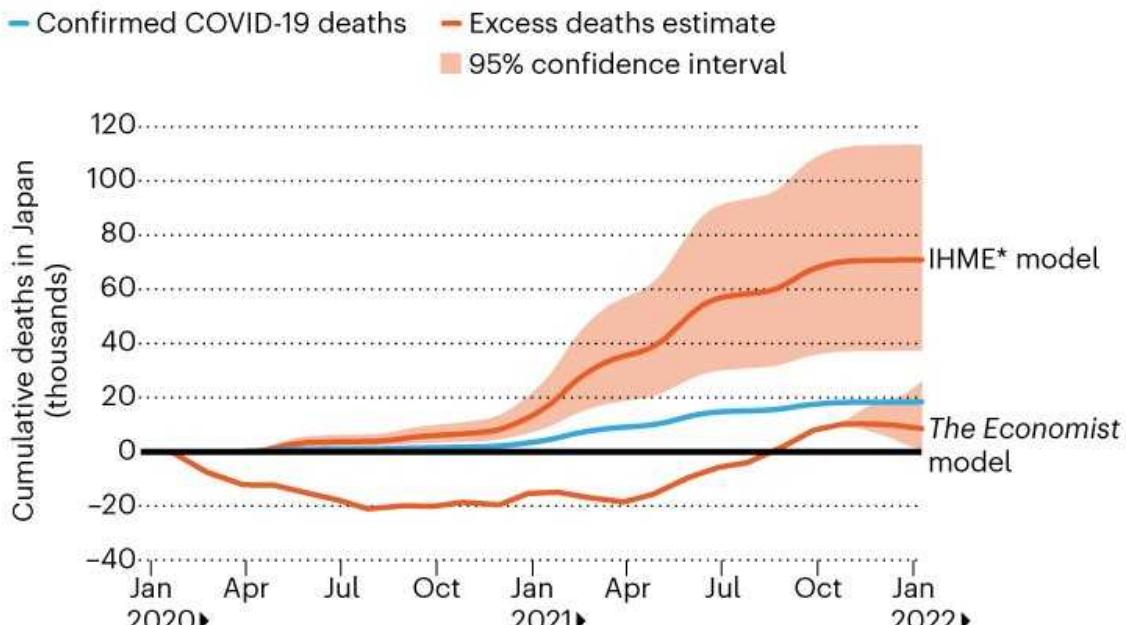
Separate estimates of real-time global deaths from the pandemic are also produced by the Institute for Health Metrics and Evaluation (IHME), an

independent global health-research centre at the University of Washington. The IHME's modelling says between 9 million and 18 million people have died so far; it also tries to forecast how this number will grow, and how fast.

Although its overall global mortality figure agrees with other estimates, there are significant differences at the national level. For example, the IHME puts cumulative excess deaths at almost 71,000 for Japan, compared with the official 18,000 reported. Yet *The Economist*'s model estimates Japan's excess deaths at between 550 and 27,000 (see 'Model disagreements').

MODEL DISAGREEMENTS

Models of excess deaths sometimes sharply disagree: the IHME* estimates 71,000 excess deaths for Japan, whereas *The Economist* model puts them at below 10,000. (In this chart, periods of negative excess mortality indicate that overall deaths were below average, even as COVID-19 deaths were being recorded.)



©nature

*Institute for Health Metrics and Evaluation.

Sources: Our World in Data/*The Economist*/IHME

There are other discrepancies, too. In May, the IHME made headlines and drew criticism for suggesting that US excess deaths in the pandemic up to that time were as high as 900,000 people. That was some 300,000 greater than other estimates, such as from the US Centers for Disease Control and

Prevention and the WMD. In October, the IHME quietly reduced the May figure to 670,000 after making changes to its modelling strategy, which some in the field complain is opaque and hard to follow.

The IHME says it will soon publish a paper detailing its model. It also says its initial US excess-death estimate was too high because it had not taken into account that winter deaths from influenza and respiratory syncytial virus might fall, and that it could include this information only once official data came in months later.

Better estimates

Even the best models are only as good as the data they rest on. Through the WHO project, demographers and others are searching for ways to improve counts and estimates of death tolls in countries that don't have reliable national mortality data. Researchers have shown this can be estimated, for example, by extrapolating from smaller regions in a country, where limited data might be available.

In a study⁷ that has not yet been peer reviewed, Karlinsky used deaths reported in a regional newspaper for the Argentinian province of Córdoba to extrapolate a nationwide excess-death estimate of 120,155 from March 2020 to August 2021, compared with official COVID-19 deaths for the period of 111,383.

Another method is to survey a representative sample of households to ask them about deaths. “This is essentially how annual number of deaths are estimated in countries without good vital registration, like Bangladesh,” Karlinsky says. Such surveys are under way in many countries and, in some cases, have already shown that excess mortality is several times higher than official COVID-19 deaths.



Crematorium workers in Bangalore in April 2021. COVID-19 deaths in India are estimated to be many times higher than official statistics. Credit: Abhishek Chinnappa/Getty

This month, for instance, a team led by epidemiologist Prabhat Jha at the University of Toronto in Canada reported the results of a telephone survey of adults in India conducted by a private polling agency tracking the pandemic. The team found that there were more than 3 million COVID-19 deaths in India up to July 2021, an estimate backed up by examining mortality data in health facilities and civil-registration deaths in ten states. The researchers — who note that other scientists have come to similar conclusions — estimate that, as of September 2021, India's COVID-19 deaths were 6–7 times higher than official statistics⁵.

Mervat Alhaffar, a public-health researcher at the London School of Hygiene and Tropical Medicine (LSHTM), worked on a study that used an even more direct method to estimate deaths: counting graves. Using satellite images of 11 cemeteries in Aden province in Yemen, the study suggested that weekly burials increased by up to 230% between April and September 2020. It estimated that, as a result of the COVID-19 pandemic, excess deaths

for the region were 2,120 during the same period⁸. Another LSHTM team has applied the same technique to count fresh graves in Mogadishu, Somalia, estimating⁹ that the city's excess death toll between January and September 2020 was 3,200 to 11,800.



An aerial view of a cemetery in Aden, Yemen, where dozens of fresh graves appeared in May 2020 as COVID-19 spread through the region. Credit: AP/Shutterstock

Alhaffar says the technique is useful, but can't be applied everywhere. "You need to engage with the locals on the ground, to understand the burial practices and make sense of the images," she says. It can be hard to establish such connections, she adds, because people in conflict zones often fear the reaction of local authorities.

And, in countries where data are scarce, cultural burial practices are harder to track. "In some places, where people might prefer to bury their loved ones in smaller graveyards nearer to their houses rather than in the big ones, analysing satellite images of cemeteries can be much more challenging," Alhaffar says.

Amid the search for ways to count deaths, Andrew Noymer, a demographer at the University of California, Irvine, says the pandemic and the increased demand for real-time mortality figures highlight a demographic shortcoming that goes back decades: many countries simply don't collect good data on births, deaths and other vital statistics. "Demographers have been part of the problem, because we have helped to put band-aids on this for 60 years. We've developed all sorts of techniques to estimate demographic rates in the absence of hard data," he says.

That means the true death toll of COVID-19 might always be disputed. "We still don't know how many people died in the 1918 [flu] pandemic, but I always figured we would know pretty well how many people would die in the next one, because we live in the modern world," Noymer says. "But we don't actually, and that's kind of sad for me as a demographer."

Nature **601**, 312-315 (2022)

doi: <https://doi.org/10.1038/d41586-022-00104-8>

References

1. 1.

De Nicola, G., Kauermann, G. & Hohle, M. *AStA Wirtsch. Sozialstat. Arch.* <https://doi.org/10.1007/s11943-021-00297-w> (2022).

2. 2.

Karlinsky, A. & Kobak, D. *eLife* **10**, e69336 (2021).

3. 3.

Simonsen, L. *et al.* *PLoS Med.* **10**, e1001558 (2013).

4. 4.

Paget, J. *et al.* *J. Glob. Health* **9**, 020421 (2019).

5. 5.

Jha, P. *et al. Science* <https://doi.org/10.1126/science.abm5154> (2022).

6. 6.

Anand, A., Sandefur, J. & Subramanian, A. *CGD Working Paper 589* (Center for Global Development, 2021).

7. 7.

Karlinsky, A. Preprint at medRxiv
<https://doi.org/10.1101/2021.08.30.21262814> (2021).

8. 8.

Koum Besson, E. S. *et al. BMJ Glob. Health* **6**, e004564 (2021).

9. 9.

Warsame, A. *et al. Int. J. Infect. Dis.* **113**, 190–199 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00104-8>

Opinion

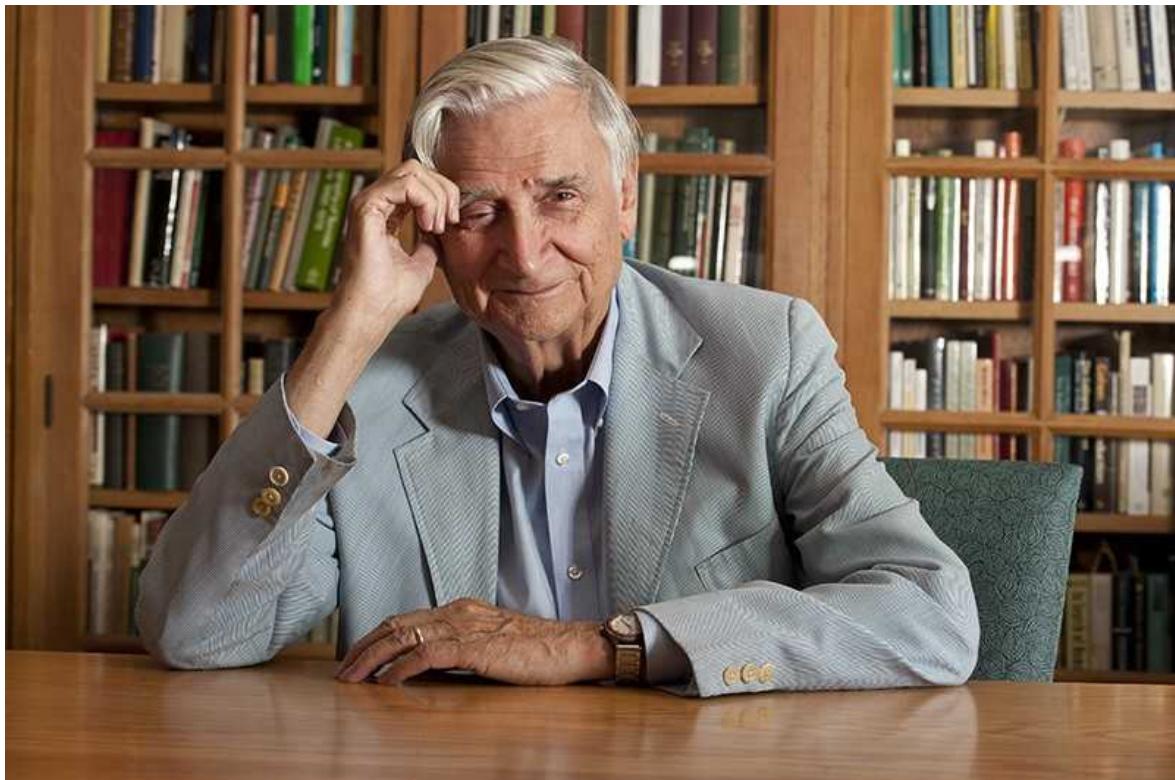
- **[Edward O. Wilson \(1929–2021\)](#)** [10 January 2022]
Obituary • Naturalist, conservationist and synthesizer who founded sociobiology.
- **[Countries should boycott Brazil over export-driven deforestation](#)** [18 January 2022]
Correspondence •
- **[Clinician-scientists: how to reverse their declining numbers](#)**
[18 January 2022]
Correspondence •
- **[Put defence money into planetary emergencies, urge Nobel winners](#)** [18 January 2022]
Correspondence •
- **[Portugal leads with Europe's largest marine reserve](#)** [18 January 2022]
Correspondence •

- OBITUARY
- 10 January 2022
- Correction [12 January 2022](#)

Edward O. Wilson (1929–2021)

Naturalist, conservationist and synthesizer who founded sociobiology.

- [Bert Hölldobler](#) ⁰



Harvard University Professor E.O. Wilson in his office at Harvard University in Cambridge, MA. USA Credit: Rick Friedman/Corbis via Getty

Edward (Ed) Wilson began by exploring the systematics, geographical distribution, social organization and evolution of ants. He became one of the great scholarly synthesizers, winning two Pulitzer prizes. A superb naturalist

who enjoyed challenging dogma, he fought for conservation, brought ideas of biodiversity into the mainstream and set ecology on a rigorous conceptual footing. He has died aged 92.

Wilson's book *Sociobiology*, published in 1975, was the first to address the evolution and organization of societies in organisms ranging from colonial bacteria to primates, including humans. The final chapter, on human social interaction, ignited controversy. Wilson argued that human behaviour, although adaptable to environmental conditions, is rooted in a genetic 'blueprint'. Opponents claimed that nothing in human behaviour is grounded in genetics, except sleeping, eating and defecation. In a letter to *The New York Review of Books*, a group of academics including evolutionary biologists Stephen Jay Gould and Richard Lewontin associated Wilson's view with racism and genocide. Wilson responded with elegance and humour; in my view, most scholars now agree that he won this argument.

Wilson was born in 1929 in Birmingham, Alabama, and grew up, as he admitted in his 1994 autobiography, *Naturalist*, "mostly insulated from its social problems". After studying biology at the University of Alabama in Tuscaloosa, he did graduate studies at Harvard University in Cambridge, Massachusetts. He felt its Museum of Comparative Zoology, with the world's largest ant collection, was his "destiny".

In 1955, he obtained his PhD on the systematics of the ant genus *Lasius*, which includes the widespread black garden ant. Systematic biology and the study of biodiversity remained his mission, but he made significant contributions to other fields, such as animal behaviour and chemical ecology. His early work on chemical communication in animals, particularly social insects, inspired a generation of scientists to explore a new area in behavioural physiology.

In 1954, Wilson set out for Melanesia, including New Guinea, to study ant taxonomy and biogeography. On the basis of his data, he elaborated the critique that he and his Harvard colleague William Brown had previously developed on the idea of subspecies. They argued that the distinctions between species should be more clearly defined, allowing for variability within species. Equally influential was their thinking on character

displacement — when similar species in the same area diverge genetically to avoid competing for resources.

Through his fieldwork in Melanesia and later in the Caribbean, Wilson drafted a principle of biogeography that he called the taxon cycle. Species evolve back and forth between being able to live in marginal habitats, and thus disperse widely, and restricting their distribution to species-rich habitats in island interiors. He tested this and other original hypotheses in the Florida Keys in the 1960s, in collaboration with his former student Daniel Simberloff. With ecologist Robert MacArthur, he proposed that species maintain their populations through trade-offs between number of offspring and quality of parental care (the concept of *r/K* selection). Their 1967 book *The Theory of Island Biogeography* had far-reaching effects on studies of evolution and conservation.

From early in his career, Wilson wondered about ways to understand the evolution of social organization, from primates to social insects (such as honeybees and ants). “A congenital synthesizer,” he wrote in his autobiography, “I held on to the dream of a unifying theory.” He developed a theory of adaptive demography — that certain kinds of social structure might increase reproductive fitness — and the evolution of division of labour between castes, such as insect queens and worker groups. First brought together in *The Insect Societies* (1971), these concepts were elaborated in *Caste and Ecology in the Social Insects*, with mathematical biologist George Oster, in 1978.

Sociobiology was a much more far-reaching synthesis on the evolution of social systems. The furore that ensued stimulated Wilson to write an even more provocative book, *On Human Nature* (1978). This garnered his first Pulitzer. His highly original book *Biophilia* (1984) was the first to use the term to mean human empathy for the natural world. He argued that pleasure in being surrounded by diverse living organisms is a biological adaptation. These books prepared the ground for *Consilience* (1998), which one reviewer called a biologist’s dream of the unity of knowledge. It proposed the kind of intellectual annexation that occurs when one field can be explained in terms of a more fundamental discipline, and received a mixed response.

To his and my utmost surprise, in 1990, the huge monograph *The Ants*, on which we worked for years, won another Pulitzer. Wilson continued to publish on human evolution and humanity's relationship with the planet into his 90s. *Half-Earth* (2016) is a passionate plea to leave half of our world to nature.

Ed was not a team builder. He preferred to work alone, although in a few cases he found colleagues who complemented his abilities. He thrived on controversy. In the past two decades, he had rejected the theory of inclusive fitness — the idea that the reproductive success of an individual increases when it helps to raise the offspring of its close relatives — that he once propagated. This led to heated debates, and I opposed some of his views. When we reached a compromise and submitted the manuscript of our book *The Superorganism* (2009), Ed's concluding remark was: "Bert, there is one thing we agree on 100%. That is: my co-author is wrong." One could disagree with Ed over scientific issues and remain good friends.

Nature **601**, 317 (2022)

doi: <https://doi.org/10.1038/d41586-022-00078-7>

Updates & Corrections

- **Correction 12 January 2022:** An earlier version of this obituary erroneously gave the date of publication of the autobiography *Naturalist* as 2006.

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00078-7>

- CORRESPONDENCE
- 18 January 2022

Countries should boycott Brazil over export-driven deforestation

- [Lucas Ferrante](#) ⁰ &
- [Philip M. Fearnside](#) ¹

Deforestation in Brazil is now threatening the last major block of intact Amazon rainforest. We urge countries that import large amounts of Brazilian soya beans and beef to take action to halt this destruction. About 9.5% of Brazil's exported soya beans and 5.3% of its beef goes to Europe, with another 79% and 52%, respectively, going to China ([P. M. Fearnside *Science* 373, 1209–1210; 2021](#)). In our view, importers should switch to other sources until Brazil eliminates export-driven deforestation.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 318 (2022)

doi: <https://doi.org/10.1038/d41586-022-00094-7>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00094-7>

- CORRESPONDENCE
- 18 January 2022

Clinician-scientists: how to reverse their declining numbers

- [Farah R. W. Kools](#) ORCID: <http://orcid.org/0000-0002-0921-0579>⁰
&
- [Berent Prakken](#) ORCID: <http://orcid.org/0000-0001-8488-4816>¹

The COVID-19 pandemic has underscored the immense importance of clinician-scientists. Central to translational medicine, they are expected to bridge the ‘valley of death’ — the gap between bench and bedside. But clinician-scientists have faced difficult career paths for decades. They must juggle demanding duties in patient care with being judged mostly on grant, publication and citation records. The result? A noticeable decline in the number of clinician-scientists ([K. Noble *et al. Nature Cancer* 1, 139–141; 2020](#)).

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 318 (2022)

doi: <https://doi.org/10.1038/d41586-022-00095-6>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00095-6>

- CORRESPONDENCE
- 18 January 2022

Put defence money into planetary emergencies, urge Nobel winners

- [Matteo Smerlak](#) ⁰ &
- [Carlo Rovelli](#) ¹

The Global Peace Dividend initiative was launched last month by more than 50 Nobel laureates and the presidents of 5 major science academies. It calls on all countries to jointly reduce military spending by 2% each year and instead contribute to a global fund to tackle climate change, pandemics and extreme poverty. We, the organizers of the initiative, encourage more people to sign the petition here: <https://peace-dividend.org/>.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 318 (2022)

doi: <https://doi.org/10.1038/d41586-022-00096-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00096-5>

- CORRESPONDENCE
- 18 January 2022

Portugal leads with Europe's largest marine reserve

- [Filipe Alves](#) ORCID: <http://orcid.org/0000-0003-3752-2745>⁰,
- [João G. Monteiro](#) ORCID: <http://orcid.org/0000-0002-3401-6495>¹,
- [Paulo Oliveira](#)² &
- [João Canning-Clode](#) ORCID: <http://orcid.org/0000-0003-2143-6535>³

Marine conservation is central to the United Nations' Sustainable Development Goals 13 (climate action) and 14 (life below water). Portugal has now created the largest marine reserve with full protection in Europe and the North Atlantic, an achievement that other nations could follow.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Nature **601**, 318 (2022)

doi: <https://doi.org/10.1038/d41586-022-00093-8>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00093-8>

Work

- **Why Joe Biden's bid to restore scientific integrity matters**

[17 January 2022]

Career Feature • Federal whistle-blowers share stories about political interference in science, and explain why the long-awaited measures announced last week are needed.

- **Preserving the legacy of Indigenous tattoos**

[17 January 2022]

Where I Work • Anthropologist Lars Krutak finds meaning in markings that are more than skin-deep.

- CAREER FEATURE
- 17 January 2022

Why Joe Biden's bid to restore scientific integrity matters

Federal whistle-blowers share stories about political interference in science, and explain why the long-awaited measures announced last week are needed.

- [Virginia Gewin](#) ⁰

[Find a new job](#)



Federal scientist Linda Birnbaum says she was sometimes blocked from speaking to the press.

In September 2019, then-president Donald Trump falsely stated that Alabama was under threat from Hurricane Dorian as it approached the US mainland.

Three days later, despite assurances from local weather bureau officials that the claim was false, Trump showed reporters a map in which the storm's projected path seemed to have been altered with a Sharpie permanent marker. The National Oceanic and Atmospheric Administration (NOAA), a federal agency, endorsed Trump's assertion.

In June 2020, a NOAA review panel found that Neil Jacobs, an atmospheric scientist and the agency's acting administrator, and Julie Roberts, its deputy chief of staff and communications director, had "engaged in misconduct intentionally, knowingly or in reckless disregard" for the agency's scientific-integrity policy by backing Trump's incorrect assertion.

The incident, dubbed Sharpiegate, features in 'Protecting the Integrity of Government Science', a long-awaited report that the Biden administration's Task Force on Scientific Integrity released last week (see go.nature.com/3ztsjv6; see also *Nature* **601**, 310–311; 2022). Ordered by the current US president seven days after his inauguration in January last year, the task force's review of scientific-integrity policies at federal agencies sets out how trust in government can be restored through scientific integrity and evidence-based policymaking.

The report calls for an overarching body that works across federal government agencies to ensure and promote best practices, and to tackle scientific-integrity violations by senior officials that cannot be handled at the agency level. These include political interference and suppression or distortion of data.

According to the Silencing Science Tracker, Sharpiegate is one of some 500 documented attempts to restrict, prohibit or misuse scientific research, education or discussion since Trump's election win in November 2016. The tracker is a joint initiative of the Climate Science Legal Defense Fund (a

non-profit organization that assists climate scientists who are silenced or face legal action because of their findings or fields of study) and Columbia University's Sabin Center for Climate Change Law, both based in New York City.

Concerns about political interference have plagued many US administrations, says Lauren Kurtz, the Climate Science Legal Defense Fund's executive director. Kurtz says that 2021 — when the fund helped 41 scientists with legal issues including censorship, open records and scientific advocacy — was one of its busiest years. A 2018 analysis found that scientific-integrity violations in the US government had increased in breadth and number in recent years (see go.nature.com/34pcjxv). Most agencies don't have robust scientific-integrity policies, she adds; nor are the policies they do have applied consistently among employees and the contractors who are hired to work on specific government projects.

The task force's 67 members, appointed from 29 government agencies, reviewed agency policies, plus comments and suggestions from more than 200 individuals and organizations. Its focus was on how to prevent political interference in the communication of science, and how to improve the transparency of scientific-integrity policies. Also, more than 650 individuals attended 3 listening sessions last summer — indicating, says Kurtz, “the wide variety of groups that care deeply about the issues”. She commends the report for identifying problems and potential solutions, but points to some shortcomings. For example, although the report says that agencies should establish clear consequences for scientific-integrity violations, it contains “no mention of what sorts of consequences might be considered, how those might be applied, or even what is the goal of having consequences”.

Escalating attacks

Boosting scientific integrity and transparency will take a concerted effort from all researchers — both inside and outside federal agencies.

In October 2019, after 40 years as a federal scientist, toxicologist Linda Birnbaum retired from her post as a director of the National Institute of Environmental Health Sciences (NIEHS) in Durham, North Carolina, part of

the National Institutes of Health. She says that politically motivated assaults on scientific research and findings reached new depths in March 2020.

“The overt attacks on science clearly came to a head with COVID-19,” Birnbaum says, citing as an example the tight controls placed on what officials from the Centers for Disease Control and Prevention (CDC) could and could not say about it.

During the Trump administration, Birnbaum says, she was discouraged, even blocked, from speaking to the press. Previously, she asked for clearance, but that never posed a problem, she says. It was standard procedure for her to get an internal policy review of manuscripts before submitting them to a journal for publication, and to get clearance from agency communications teams for interviews with reporters. But under Trump, she says, everything she wrote, even slides for conference presentations, had to be cleared.

Birnbaum says that she would make changes requested by deputy administrators, but that their interference could go only so far. “I would say whatever the hell I wanted to during my talk,” she says, and she told reporters that if they came up to her afterwards, she wouldn’t need clearance to speak to them. “There are certain workarounds, but it takes a lot of mental and emotional energy,” she says. Political interference in the work of federal scientists “violates the trust that the public places in government to best service its collective interests”, says the task force report. It warns that federal agencies in which scientific integrity is not protected will struggle to recruit and retrain scientists. And it distinguishes between a supervisor’s edits to a scientific report, prompted by valid concerns about analytical techniques, and the distortion of outcomes to meet preferred policy objectives, which it describes as interference.

The report recommends mandatory scientific-integrity training for federal-agency staff whose roles require them to use science to make policy decisions. And a raft of proposed measures on building trust, openness and transparency between scientists and communications staff include media training for federal scientists, and a recommendation that a scientist whose research features in a press release should enjoy the “right of last review” of scientific content.

Integrity complaints

In September 2017, marine biologist Peter Corkeron and his colleagues published data on a significant decline in numbers of the North Atlantic right whale¹. But he alleges that his NOAA superiors ignored his findings until they were published, despite his repeated warnings about the severity of the situation, dating back to February 2016. North Atlantic right whale (*Eubalaena glacialis*) recovery had been a point of pride at NOAA, Corkeron says. But once the agency anticipated being sued by conservation and animal-protection groups for failing to prevent whale numbers decreasing, he declined to fall in line with the decision by NOAA's National Marine Fisheries Service to stop putting information on the whales' status in e-mails or internal memos. In March 2018, after e-mailing his frustrations that the fisheries services was doing too little to protect the whales from extinction, Corkeron received a formal letter of reprimand to caution him that "professionalism and courtesy will be expected in all further communications".



Microbiologist Evi Emmenegger faced dismissal after raising concerns about waste water.Credit: Blake Feist

One option open to him was to file a formal scientific-integrity complaint, he says. Such complaints can include allegations of mismanagement, misconduct, abuse of authority, or censorship that compromises an agency's scientific record. At NOAA, this involves a written allegation including the names, facts, documents, witnesses and an explanation of the alleged misconduct; the complainant can remain anonymous. For various reasons, including that his complaint involved members of the body overseeing the process, Corkeron sought a new job, and since October 2019 he has led the whale-research programme at the New England Aquarium in Boston, Massachusetts.

“All you can do is decide what your ‘red line’ is,” he says, and understand that pushing back when that line is crossed will be a stressful experience. *Nature* asked NOAA to comment on Corkeron’s claims about the lack of support for his research, his letter of reprimand and the decision to avoid discussing whale declines with him in writing. A spokesperson said the agency did not comment on personnel matters, but confirmed that Corkeron and his North Atlantic right-whale research were never the subject of a scientific-integrity allegation at the agency.

In his submission to the task force, Corkeron points to a loss of public trust in federal science when large numbers of staff leave in a short time frame. He also called for an office overseeing integrity across all federal government agencies, and highlighted problems with the internal policy-review process.

Whistle-blower insights

Evi Emmenegger spent 28 years as a research microbiologist at the Western Fisheries Research Center in Seattle, Washington, a biosafety laboratory operated by the US Geological Survey (USGS). It was home to numerous pathogens, including exotic and invasive viruses. Emmenegger says that, in July 2017, she raised concerns to her supervisors about contaminated waste water being released in nearby wetlands over a six-month period, but that no

immediate corrective actions were taken. She filed a scientific-integrity complaint in mid-September, according to Jeff Ruch, a director at Public Employees for Environmental Responsibility (PEER), a non-profit organization in Silver Spring, Maryland, that supports potential government whistle-blowers. In February 2020, PEER highlighted Emmenegger's case on its website. It claimed that the USGS had dismissed her complaint and sought to fire her because of alleged issues with the quality of a research paper that she had prepared. She was placed on administrative leave for 13 months, but was officially reinstated on 10 May 2021.

Before her reinstatement, Emmenegger told *Nature* she thought that the US Whistleblower Protection Act — passed in 1978 and updated in November 2012 to permit whistle-blowers to collect compensatory damages — lacks teeth.

The current system, she said, is concerned mainly with damage control and does little to protect those who reveal alleged wrongdoings. The task force report urges agencies to post its scientific-integrity policies online, along with instructions for reporting concerns, and to publish regular public reports on integrity violations and how they are addressed.



Staff who care about workplace ethics are valuable, says Joel Clement (pictured second from right). Credit: Benn Craig/Belfer Center

Emmenegger also advised whistle-blowers to hire a lawyer and to contact government-accountability groups. In the United States, these include PEER; the Union of Concerned Scientists (UCS), a science-advocacy non-profit organization based in Cambridge, Massachusetts; and the Environmental Protection Network, a group of more than 550 former Environmental Protection Agency (EPA) staff and political employees formed in January 2017. The task force report stresses the importance of protecting those who report alleged scientific-integrity violations.

Ruch is representing Emmenegger in negotiations with the USGS over her return to work; the agency declined to comment for this story. PEER describes the task force report as “underwhelming” and lacking specifics, “all but ensuring the Biden administration will fall short on its effort to strengthen federal scientific-integrity policies”.

Employability

Kurtz advises people who are considering filing a formal complaint to “see if the scientific-integrity office of their agency has put any reports out, to see how previous complaints have been resolved”. It’s also a good idea to gauge the level of internal support for a complaint. “Still, the best option is to go to the [agency’s] scientific-integrity official,” says Jacob Carter, a senior scientist at the UCS. Sometimes, that person will schedule a meeting off the record to explore whether a formal complaint has merit, he adds.

In 2017, Joel Clement, a former climate-policy adviser at the Department of the Interior, filed a whistle-blower complaint after being moved to an accounting post from his role helping Alaska Native communities to adapt to climate change. Clement claims that the move was triggered after he spoke out publicly — including at a United Nations conference — about the dangers these communities face from a changing climate. He described what happened to him in an opinion piece in *The Washington Post*. *Nature* asked the Department of the Interior to comment on why Clement was moved to an accounting post, but it declined.

Clement says that some organizations — including his current employer, the Belfer Center for Science and International Affairs at the Harvard Kennedy School in Cambridge — will hire former whistle-blowers. “You don’t have to wear a scarlet W [for whistle-blower] — showing that you care about ethics of the workplace is valuable to employers,” he says.

Soon after Trump’s election, about ten environmental researchers in academia and non-profit organizations brainstormed ways to monitor government data integrity and website information as a pre-emptive strike against possible data loss. The result was the formation of the Environmental Data & Governance Initiative (EDGI), which archived government data and also monitored federal websites and created a new academic field, environmental data justice.



Leif Fredrickson told Congress about changes to the enforcement of environmental regulations. Credit: Leif Fredrickson

“We were very concerned about the possibility of data loss,” says the EDGI’s co-founder Sara Wylie, a researcher studying large-scale environmental-health issues at Northeastern University in Boston.

“What we did see was more subtle than that; it was the removal of key data and background information — the biggest shift being the removal of the

EPA's climate-change page from its website in January 2017, which was a clear signal about the administration's priorities, she says.

"There is an enormous space for researchers from all different fields to gather and pool their expertise and approaches to do work in the public interest rapidly."

Stony Brook University in Long Island, New York, hosts the EDGI's mirror of an EPA database whose records track the agency's enforcement of federal environmental laws. This mirroring makes the information more accessible and understandable to the public. "We are actively building partnerships with academics and NGOs to use these tools to share EPA enforcement data in meaningful ways to engage the public," Wylie adds.

The team also interviewed 50 long-term federal-agency employees and wrote 'The First 100 Days and Counting' (see go.nature.com/3tmxa2f). This report documented the fossil-fuel industry's influence on the Trump administration, changes in how climate science was presented to the public and the administration's hostility to scientific research and evidence.

The researchers followed that up with two peer-reviewed articles — the first looking at the EPA early in the Trump administration²; the second recounting US presidential interventions on environmental health protection³. They also produced a report entitled *A Sheep in the Closet*⁴. The authors testified before Congress about declines in the Trump administration's enforcement of EPA regulations.

The Wayback Machine Internet archive has implemented EDGI-developed software to let others track changes to websites, says Wylie. "The key element is to make it possible for others to build on and iterate on [the software], with a clear set of guiding values that are shared publicly," she says, adding that a high level of transparency is crucial for organizations aiming to hold federal agencies accountable.

Leif Fredrickson, an environmental-policy historian at the University of Montana in Missoula and an EDGI member, says that the convoluted presentation of government information obscures the EPA's overall mission.

The EDGI has therefore created what it calls the People's EPA, an attempt to make the agency's data and policy clearer for the public.

"We're going to put up long-term data sets on stuff like budget and staffing for EPA so that when budget numbers come out, they can be quickly contextualized in a broader history," says Fredrickson, who has interviewed 90 former and current EPA staffers. "It's going to take a lot of different organizations to move in this direction" to ensure a better public understanding of government science.

Nature **601**, 471–473 (2022)

doi: <https://doi.org/10.1038/d41586-022-00105-7>

References

1. 1.

Pace, R. M. III, Corkeron, P. J. & Kraus, S. D. *Ecol. Evol.* **7**, 8730–8741 (2017).

2. 2.

Dillon, L. *et al.* *Am. J. Public Health* **108**, S89–S94 (2018).

3. 3.

Fredrickson, L. *et al.* *Am. J. Public Health* **108**, S95–S103 (2018).

4. 4.

Fredrickson, L. *et al.* *A Sheep in the Closet: the Erosion of Enforcement at the EPA* (Environmental Data & Governance Initiative, 2019).

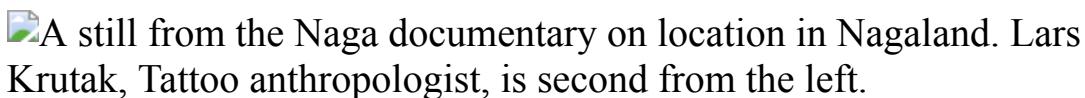
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00105-7>

- WHERE I WORK
- 17 January 2022

Preserving the legacy of Indigenous tattoos

Anthropologist Lars Krutak finds meaning in markings that are more than skin-deep.

- [Chris Woolston](#) ⁰



Lars Krutak is an anthropologist with the Museum of International Folk Art in Santa Fe, New Mexico. Credit: Michael Zomer

Tattoos tell stories and connect people to their pasts. In this picture taken in November 2019, I take notes as Chen-o Khuzuthrupa, a centenarian nobleman of the Chen tribe of northeastern India, describes a tattoo on his back: a sacred tiger familiar spirit. He received the tattoo — a tiger stripe motif with circles that represent the spirit — during an elaborate ceremony in the 1930s, when he was a young headhunter and warrior. He says the markings gave him a tiger's strength in battle and the ability to spy on his enemies in his dreams. Those powers are gone. He lost the connection to the tiger spirit decades ago when he converted to Christianity.

The picture is a still from [*Patterns of Life*](#), a documentary of Indigenous tattoos that is currently available on Instagram. As a tattoo anthropologist, I work to document, preserve and interpret a neglected part of cultural heritage. Young people in these villages often sport modern tattoo designs, such as barbed wire across the biceps. I've asked them why they prefer images from other cultures, and they explain that they want to move

forward, not look back. And modern tattoos have a lot more colour. Khuzuthrupa's tattoos were made from soot from a fireplace.

Young people can put whatever they want on their bodies, but they should at least be aware of the designs of their ancestors because it's a part of their heritage. That's why I wrote *Kalinga Tattoo*, a 2010 book about tattoos of the northern Philippines, and sent a pallet of copies to villages in that area. Those images shouldn't just sit around in my office.

Anyone could get a tiger spirit tattoo in the style of Khuzuthrupa's, but much of its significance would be lost without the culture and context around it. The original artist is long dead, and nobody remembers the prayers, songs and other aspects of the tattooing ceremony. The two other men in the village with this particular tattoo are in their nineties. They are the last generation to carry the tiger spirit.

Nature **601**, 476 (2022)

doi: <https://doi.org/10.1038/d41586-022-00106-6>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00106-6>

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

Research

- **[A blood test to predict complications of pregnancy](#)** [05 January 2022]
News & Views • Cell-free RNA transcripts in maternal blood can be analysed to monitor the progression of pregnancy and to predict a potentially harmful pregnancy-specific condition called pre-eclampsia.
- **[Silicon qubits move a step closer to achieving error correction](#)** [19 January 2022]
News & Views • A silicon-based quantum-computing platform has met key standards for reducing error — setting the stage for quantum devices that could benefit from established semiconductor microchip technologies.
- **[A virtual drug-screening approach to conquer huge chemical libraries](#)** [15 December 2021]
News & Views • A computational method has been devised to identify drug-candidate molecules from a library of billions of molecules using 100 times less computational power than is used by standard methods.
- **[Nanoparticle asymmetry shapes an immune response](#)** [19 January 2022]
News & Views • The chirality, or handedness, of nanoparticles is shown to be a key factor in determining how well such particles engage with the immune system — a finding that might help to inform the design of vaccines and anticancer therapeutics.
- **[Future ice loss captured by historical snapshots](#)** [19 January 2022]
News & Views • Archival images of glacial ice on a Norwegian archipelago, together with the islands' climatic diversity, enable application of an innovative method for making long-term projections of ice loss using short-term observations.
- **[A radiocarbon revolution sheds light on the Vikings](#)** [22 December 2021]
News & Views • Advances in the precision of radiocarbon dating can offer year-specific data. Analyses of archaeological sites in Denmark and Canada provide insights into the chronology of the global networks of the Viking Age.
- **[Black-hole-triggered star formation in the dwarf galaxy Henize 2-10](#)** [19 January 2022]
Article • Optical observations with a linear resolution of a few parsecs show that the outflow from the central black hole in the low-mass galaxy Henize 2-10 triggered a round of star

formation.

- **Star formation near the Sun is driven by expansion of the Local Bubble** [12 January 2022]

Article • Three-dimensional analysis of the solar neighbourhood shows that nearly all star-forming regions near the Sun lie on the surface of the Local Bubble, which was inflated by supernovae about 14 million years ago.

- **Fast universal quantum gate above the fault-tolerance threshold in silicon** [19 January 2022]

Article • Single- and two-qubit gate fidelities above the fault-tolerance threshold for quantum computation are demonstrated in silicon quantum dots by fast electrical control using a micromagnet-induced gradient field and tunable coupling.

- **Quantum logic with spin qubits crossing the surface code threshold** [19 January 2022]

Article • A spin-based quantum processor in silicon achieves single-qubit and two-qubit gate fidelities above 99.5% using gate-set tomography, exceeding the theoretical threshold required for fault-tolerant quantum computing.

- **Precision tomography of a three-qubit donor quantum processor in silicon** [19 January 2022]

Article • Universal quantum logic operations with fidelity exceeding 99%, approaching the threshold of fault tolerance, are realized in a scalable silicon device comprising an electron and two phosphorus nuclei, and a fidelity of 92.5% is obtained for a three-qubit entangled state.

- **Topological triple phase transition in non-Hermitian Floquet quasicrystals** [19 January 2022]

Article • A triple phase transition, where changing a single parameter simultaneously gives rise to metal–insulator, topological and a parity–time symmetry-breaking phase transitions, is observed in non-Hermitian Floquet quasicrystals.

- **Chemical crystallography by serial femtosecond X-ray diffraction** [19 January 2022]

Article • Small-molecule serial femtosecond X-ray crystallography (smSFX) characterizes microcrystals by indexing sparse serial XFEL diffraction frames, with little sample preparation, without beam damage, and at room temperature and pressure.

- **Enantiomer-dependent immunological response to chiral nanoparticles** [19 January 2022]

Article • Nanoparticles with different chiralities have different in vitro and in vivo effects on the immune system, suggesting new ways of creating vaccine adjuvants.

- **Historical glacier change on Svalbard predicts doubling of mass loss by 2100** [19 January 2022]

Article • Historical photographs, modern observations, and a simple space-for-time substitution approach predict that glacier mass loss from Svalbard is poised to double over the twenty-first century.

- **Global fine-scale changes in ambient NO₂ during COVID-19 lockdowns** [19 January 2022]

Article • The satellite instrument TROPOMI is used to assess ambient NO₂ levels at approximately one-kilometre resolution across 215 cities worldwide during COVID-19 lockdowns, finding about 30% lower NO₂ concentrations in countries with strict lockdowns.

- **Evidence for European presence in the Americas in ad 1021** [20 October 2021]

Article • Precise dating of wooden artefacts at a Norse settlement in Newfoundland establishes that the Norse were in the Americas in ad 1021.

- **Single-year radiocarbon dating anchors Viking Age trade cycles in time** [22 December 2021]

Article • Disturbances in the radiocarbon record anchor a precisely dated archaeological stratigraphy of a medieval trading emporium in Denmark in time, revealing that the Viking expansion was associated with competition for trade routes rather than with raids.

- **Individual human cortical progenitors can produce excitatory and inhibitory neurons** [15 December 2021]

Article • Molecular barcoding is used to show that progenitor cells in the human cortex can produce both excitatory neurons and inhibitory interneurons, with implications for our understanding of the evolution of the human brain.

- **Single-cell delineation of lineage and genetic identity in the mouse brain** [15 December 2021]

Article • Single-cell RNA sequencing with parallel tagging of progenitor cells is used to track clonal relationships and transcriptomic signatures during development of the mouse forebrain.

- **Optimization of non-coding regions for a non-modified mRNA COVID-19 vaccine** [18 November 2021]

Article • CV2CoV, a second-generation mRNA COVID-19 vaccine with non-modified nucleosides but optimized non-coding regions, is demonstrated to be effective against SARS-CoV-2 challenge when tested in non-human primates.

- **Behavioural immune landscapes of inflammation** [05 January 2022]

Article • High-dimensional datasets derived from time-resolved live imaging of leukocytes in mice were used to identify leukocyte identities and dynamic neutrophil states with high

cellular resolution.

- **RNA profiles reveal signatures of future health and disease in pregnancy** [05 January 2022]
Article • Expression signatures from cell-free RNA of pregnant women can be used to reveal normal biology of pregnancy and predict development of pre-eclampsia.
- **Valine tRNA levels and availability regulate complex I assembly in leukaemia** [22 December 2021]
Article • Restriction of dietary valine reduces growth of T cell acute lymphoblastic leukaemia through altered valine tRNA biogenesis and reduced translation of mRNAs that encode subunits of mitochondrial complex I.
- **Targeting SWI/SNF ATPases in enhancer-addicted prostate cancer** [22 December 2021]
Article • PROTAC degrader-induced SWI/SNF inactivation abolishes DNA accessibility at enhancer elements of oncogenes and also tempers supra-physiologic expression of driver transcription factors, resulting in potent inhibition of tumour growth in mouse models.
- **Genome surveillance by HUSH-mediated silencing of intronless mobile elements** [18 November 2021]
Article • The human silencing hub (HUSH) complex uses introns to distinguish intronless foreign DNA from intron-containing host DNA and modifies chromatin to silence transcription of retrotransposons and retroviruses.
- **MicroRNA sequence codes for small extracellular vesicle release and cellular retention** [22 December 2021]
Article • MicroRNAs encode sorting sequences that determine whether they are secreted in exosomal vesicles to regulate gene expression in distant cells or retained in cells that produced them, with different sequences used by individual cell types.
- **Synthon-based ligand discovery in virtual libraries of over 11 billion compounds** [15 December 2021]
Article • V-SYNTHES, a scalable and computationally cost-effective synthon-based approach to compound screening, identified compounds with a high affinity for CB2 and CB1 in a hierarchical structure-based screen of more than 11 billion compounds.
- **Structure of Hsp90–Hsp70–Hop–GR reveals the Hsp90 client-loading mechanism** [22 December 2021]
Article • The cryo-electron microscopy structure of the glucocorticoid receptor (GR)-loading complex—a complex in which Hsp70 loads GR onto Hsp90 and Hop—is described, providing insights into how the chaperones Hsp90 and Hsp70 coordinate to facilitate GR remodelling for activation.

- **Structure of Hsp90–p23–GR reveals the Hsp90 client-remodelling mechanism** [22 December 2021]

Article • Studies based on cryo-electron microscopy structures of Hsp90 chaperone complexes reveal the molecular mechanism of the chaperone-mediated maturation of the human glucocorticoid receptor.

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

- NEWS AND VIEWS
- 05 January 2022

A blood test to predict complications of pregnancy

Cell-free RNA transcripts in maternal blood can be analysed to monitor the progression of pregnancy and to predict a potentially harmful pregnancy-specific condition called pre-eclampsia.

- [Lydia L. Shook](#)⁰ &
- [Andrea G. Edlow](#)¹

The biology of human pregnancy is challenging to study. Invasive diagnostic procedures can be used to obtain DNA from the placenta or fetus for genetic analyses, or to sample RNA for a snapshot of development at that moment in time. For ethical and practical reasons, however, it is not possible to repetitively sample the growing fetus and placenta to monitor development and pregnancy health across gestation. [Writing in Nature](#), Rasmussen *et al.*¹ collect and analyse RNA molecules circulating freely in the bloodstream — known as cell-free RNA (cfRNA) — from more than 1,800 pregnant individuals of various ages, body mass indices and races, from different continents and at different stages of their pregnancies. The analyses provide insight into normal fetal, placental and maternal changes in gene expression across gestation. The authors then use this knowledge of normal cfRNA signatures in pregnancy to predict the development of a potentially dangerous pregnancy complication called pre-eclampsia.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 319-320 (2022)

doi: <https://doi.org/10.1038/d41586-021-03801-y>

References

1. 1.
Rasmussen, M. *et al.* *Nature* **601**, 422–427 (2022).
2. 2.
Edlow, A. G. & Bianchi, D. W. *Biochim. Biophys. Acta* **1822**, 1970–1980 (2012).
3. 3.
Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L. & Quake, S. R. *Proc. Natl Acad. Sci. USA* **105**, 16266–16271 (2008).
4. 4.
Chiu, R. W. K. *et al.* *BMJ* **342**, c7401 (2011).
5. 5.
ACOG. *Obstet. Gynecol.* **135**, e237–e260 (2020).
6. 6.
Stepan, H., Hund, M. & Andraczek, T. *Hypertension* **75**, 918–926 (2020).
7. 7.
Ngo, T. T. M. *et al.* *Science* **360**, 1133–1136 (2018).
8. 8.
Koh, W. *et al.* *Proc. Natl Acad. Sci. USA* **111**, 7361–7366 (2014).

9. 9.

Kramer, A. W., Lamale-Smith, L. M. & Winn, V. D. *Placenta* **37**, 19–25 (2016).

10. 10.

Schumann, S., Buck, V. U., Classen-Linke, I., Wennemuth, G. & Grümmer, R. *Histochem. Cell Biol.* **144**, 571–585 (2015).

11. 11.

Alazami, A. M. *et al.* *Genome Biol.* **16**, 240 (2015).

12. 12.

Cunningham, P. & McDermott, L. *J. Nutr.* **139**, 636–639 (2009).

13. 13.

Tan, M. Y. *et al.* *Ultrasound Obstet. Gynecol.* **52**, 186–195 (2018).

14. 14.

Vyas, D. A., Eisenstein, L. G. & Jones, D. S. *N. Engl. J. Med.* **383**, 874–882 (2020).

15. 15.

Henderson, J. T. *et al.* *Ann. Intern Med.* **160**, 695–703 (2014).

16. 16.

Van Doorn, R. *et al.* *PLoS ONE* **16**, e0247782 (2021).

17. 17.

Lamb, J. *Nature Rev. Cancer* **7**, 54–60 (2007).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03801-y>.

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

- NEWS AND VIEWS
- 19 January 2022

Silicon qubits move a step closer to achieving error correction

A silicon-based quantum-computing platform has met key standards for reducing error — setting the stage for quantum devices that could benefit from established semiconductor microchip technologies.

- [Ada Warren](#) ⁰ &
- [Sophia E. Economou](#) ¹

Quantum bits (qubits) that use the quantum properties of electrons in silicon devices offer enormous potential for developing compact and robust quantum computers that take advantage of the existing silicon-microchip industry. But quantum operations are subject to error, and getting error rates low enough to make quantum silicon devices feasible remains a challenge. Three papers in this issue, by [Xue et al.](#)¹, [Noiri et al.](#)² and [Mądzik et al.](#)³, report demonstrations of qubit operations in silicon devices with fidelities above the threshold of one of the most popular quantum error-correcting codes. The results suggest that these devices could be a competitive platform for scalable quantum-information processing.

Access options

Subscribe to nature+

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

\$29.99

monthly

Subscribe

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Nature **601**, 320-322 (2022)

doi: <https://doi.org/10.1038/d41586-022-00047-0>

References

1. 1.

Xue, X. *et al.* *Nature* **601**, 343–347 (2022).

2. 2.

Noiri, A. *et al.* *Nature* **601**, 338–342 (2022).

3. 3.

Mądzik, M. T. *et al.* *Nature* **601**, 348–353 (2022).

4. 4.

Kandala, A. *et al.* *Nature* **549**, 242–246 (2017).

5. 5.

Arute, F. *et al.* *Nature* **574**, 505–510 (2019).

6. 6.

Chatterjee, A. *et al.* *Nature Rev. Phys.* **3**, 157–177 (2021).

7. 7.

Loss, D. & DiVincenzo, D. P. *Phys. Rev. A* **57**, 120–126 (1998).

8. 8.

Zwerver, A. M. J. *et al.* Preprint at <https://arxiv.org/abs/2101.12650> (2021).

9. 9.

Ofek, N. *et al.* *Nature* **536**, 441–445 (2016).

10. 10.

Egan, L. *et al.* *Nature* **598**, 281–286 (2021).

11. 11.

Borjans, F., Croot, X. G., Mi, X., Gullans, M. J. & Petta, J. R. *Nature* **577**, 195–198 (2020).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00047-0>

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

- NEWS AND VIEWS
- 15 December 2021

A virtual drug-screening approach to conquer huge chemical libraries

A computational method has been devised to identify drug-candidate molecules from a library of billions of molecules using 100 times less computational power than is used by standard methods.

- [Charlotte Deane](#) ⁰ &
- [Maranga Mokaya](#) ¹

Creating new drugs is extremely difficult, and very expensive¹. There are an estimated 10^{60} drug-like molecules in chemical space — the theoretical space spanned by all possible molecules and compounds². However, only relatively few will bind to a particular target site in the body and thus be medically useful. Therefore, identifying a molecule that acts on a specific target and is also suitable for use as a drug (for example, in terms of how it travels to its target and is broken down) is a highly complex task. [Writing in Nature](#), Sadybekov *et al.*³ present an innovative approach to identifying potential drug molecules from exceptionally large chemical libraries, and demonstrate its performance in a screen for inhibitors of three target proteins.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 322-323 (2022)

doi: <https://doi.org/10.1038/d41586-021-03682-1>

References

1. 1.

Schneider, P. & Schneider, G. *J. Med. Chem.* **59**, 4077–4086 (2016).

2. 2.

Bohacek, R. S., McMardin, C. & Guida, W. C. *Med. Res. Rev.* **16**, 3–50 (1996).

3. 3.

Sadybekov, A. A. *et al.* *Nature* **601**, 452–459 (2022).

4. 4.

Ross, G. A., Morris, G. M. & Biggin, P. C. *J. Chem. Theory Comput.* **9**, 4266–4274 (2013).

5. 5.

Llanos, M. A. *et al.* *J. Chem. Inf. Model.* **61**, 3758–3770 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03682-1>

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

- NEWS AND VIEWS
- 19 January 2022

Nanoparticle asymmetry shapes an immune response

The chirality, or handedness, of nanoparticles is shown to be a key factor in determining how well such particles engage with the immune system — a finding that might help to inform the design of vaccines and anticancer therapeutics.

- [Alexander Hooftman](#)⁰ &
- [Luke A. J. O'Neill](#)¹

Shape is all-important for molecules, especially when it comes to a property known as chirality (handedness). A chiral molecule exists in two forms, called enantiomers, that are chemically identical, but are mirror images of each other, and that interact with other molecules in different ways. Chiral molecules have long been used in drug design — and some molecules can even be tailored to interact with the body in enantiomer-specific ways.

[Writing in *Nature*](#), Xu *et al.*¹ report that chirality can also be used to design nanoparticles that have identical chemical structures, but that differ in their ability to activate immune cells, owing to differences in the spatial arrangement of their atoms.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 323-325 (2022)

doi: <https://doi.org/10.1038/d41586-021-03806-7>

References

1. 1.

Xu, L. *et al. Nature* **601**, 366–373 (2022).

2. 2.

Oh, N. & Park, J-H. *Int. J. Nanomed.* **9** (Suppl. 1), 51–63 (2014).

3. 3.

Muñoz-Planillo, R. *et al. Immunity* **38**, 1142–1153 (2013).

4. 4.

Coll, R. C. *et al. Nature Med.* **21**, 248–255 (2015).

5. 5.

Ndeupen, S. *et al. iScience* **24**, 103479 (2021).

6. 6.

Shin, M. D. *et al. Nature Nanotechnol.* **15**, 646–655 (2020).

7. 7.

Lebre, F., Pedroso de Lima, M. C., Lavelle, E. C. & Borges, O. *Int. J. Pharmaceut.* **552**, 7–15 (2018).

8. 8.

Mitchell, L. A., Lauer, F. T., Burchiel, S. W. & McDonald, J. D. *Nature Nanotechnol.* **4**, 451–456 (2009).

9. 9.

Smith, D. M., Simon, J. K. & Baker, J. R. Jr *Nature Rev. Immunol.* **13**, 592–605 (2013).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03806-7>

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

- NEWS AND VIEWS
- 19 January 2022

Future ice loss captured by historical snapshots

Archival images of glacial ice on a Norwegian archipelago, together with the islands' climatic diversity, enable application of an innovative method for making long-term projections of ice loss using short-term observations.

- [Twila A. Moon](#) [ORCID: http://orcid.org/0000-0003-0968-7008](#) ⁰

Melting glaciers, ice caps and ice sheets are wreaking havoc at shorelines around the world, and there is widespread recognition of the need to understand where, when and how much ice will be lost in the future as a result of climate change. Much of what we know about Earth's existing land ice is gleaned from rich remote-sensing records, yet relatively few of these records span more than a couple of decades — time periods that are short enough to be biased by sporadic glacier behaviour. [Writing in Nature](#), Geyman *et al.*¹ use observations of glaciers on the Norwegian islands that make up Svalbard to improve projections of the ice mass that is expected to be lost in this area during the twenty-first century.

Access options

Subscribe to nature+

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

\$29.99

monthly

Subscribe

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Nature **601**, 325-326 (2022)

doi: <https://doi.org/10.1038/d41586-022-00046-1>

References

1. 1.

Geyman, E. C., van Pelt, W. J. J., Maloof, A. C., Aas, H. F. & Kohler, J. *Nature* **601**, 374–379 (2022).

2. 2.

Blois, J. L., Williams, J. W., Fitzpatrick, M. C., Jackson, S. T. & Ferrier, S. *Proc. Natl Acad. Sci. USA* **110**, 9374–9379 (2013).

3. 3.

IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) (Cambridge Univ. Press, 2013).

4. 4.

Post, E. *et al. Sci. Adv.* **5**, eaaw9883 (2019).

5. 5.

Huss, M. & Hock, R. *Front. Earth Sci.* **3**, 54 (2015).

6. 6.

Radić, V. *et al. Clim. Dyn.* **42**, 37–58 (2014).

7. 7.

Marzeion, B., Jarosch, A. H. & Hofer, M. *Cryosphere* **6**, 1295–1322 (2012).

8. 8.

Fürst, J. J. *et al. Geophys. Res. Lett.* **45**, 11760–11769 (2018).

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

- NEWS AND VIEWS
- 22 December 2021

A radiocarbon revolution sheds light on the Vikings

Advances in the precision of radiocarbon dating can offer year-specific data. Analyses of archaeological sites in Denmark and Canada provide insights into the chronology of the global networks of the Viking Age.

- [James H. Barrett](#) ORCID: <http://orcid.org/0000-0002-6683-9891> ⁰

During the twentieth century, our knowledge of the past was revolutionized by the introduction of radiocarbon dating, by the calibration of radiocarbon data to calendar dates using wood controls of known age and by advances in dating increasingly small samples. Now, another radiocarbon revolution is under way. There is a growing trend in harnessing calibration data that can be pinpointed to an individual year — measured using single rings of individual trees. For the years for which such calibration data exist, the most-recent international radiocarbon calibration curve, called IntCal20, represents a principal advance over the previously used approach (which often combined several tree rings for calibration analysis)¹. Writing in *Nature*, [Philippson et al.](#)² and [Kuitems et al.](#)³ report how this method was used to clarify the timelines of two key Viking sites.

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Nature **601**, 326-327 (2022)

doi: <https://doi.org/10.1038/d41586-021-03769-9>

References

1. 1.

Reimer, P. J. *et al.* *Radiocarbon* **62**, 725–757 (2020).

2. 2.

Philippsen, B., Feveile, C., Olsen, J. & Sindbæk, S. M. *Nature* <https://doi.org/10.1038/s41586-021-04240-5> (2021).

3. 3.

Kuitems, M. *et al.* *Nature* **601**, 388–391 (2022).

4. 4.

Koch, A., Brierley, C., Maslin, M. M. & Lewis, S. L. *Quat. Sci. Rev.* **207**, 13–36 (2019).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03769-9>

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

- Article
- [Published: 19 January 2022](#)

Black-hole-triggered star formation in the dwarf galaxy Henize 2-10

- [Zachary Schutte](#) ORCID: [orcid.org/0000-0001-7412-8988¹](https://orcid.org/0000-0001-7412-8988) &
- [Amy E. Reines](#) ORCID: [orcid.org/0000-0001-7158-614X¹](https://orcid.org/0000-0001-7158-614X)

Nature volume **601**, pages 329–333 (2022)

- 864 Accesses
- 508 Altmetric
- [Metrics details](#)

Subjects

- [Astronomy and astrophysics](#)
- [Galaxies and clusters](#)

Abstract

Black-hole-driven outflows have been observed in some dwarf galaxies with active galactic nuclei¹, and probably play a role in heating and expelling gas (thereby suppressing star formation), as they do in larger galaxies². The extent to which black-hole outflows can trigger star formation in dwarf galaxies is unclear, because work in this area has previously focused on massive galaxies and the observational evidence is scarce^{3,4,5}. Henize 2-10 is a dwarf starburst galaxy previously reported to have a central massive black hole^{6,7,8,9}, although that interpretation has

been disputed because some aspects of the observational evidence are also consistent with a supernova remnant^{10,11}. At a distance of approximately 9 Mpc, it presents an opportunity to resolve the central region and to determine if there is evidence for a black-hole outflow influencing star formation. Here we report optical observations of Henize 2-10 with a linear resolution of a few parsecs. We find an approximately 150-pc-long ionized filament connecting the region of the black hole with a site of recent star formation. Spectroscopy reveals a sinusoid-like position–velocity structure that is well described by a simple precessing bipolar outflow. We conclude that this black-hole outflow triggered the star formation.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: HST optical image of the dwarf starburst galaxy Henize 2-10.



Fig. 2: Optical spectra and ionized gas kinematics for the central region of Henize 2-10.

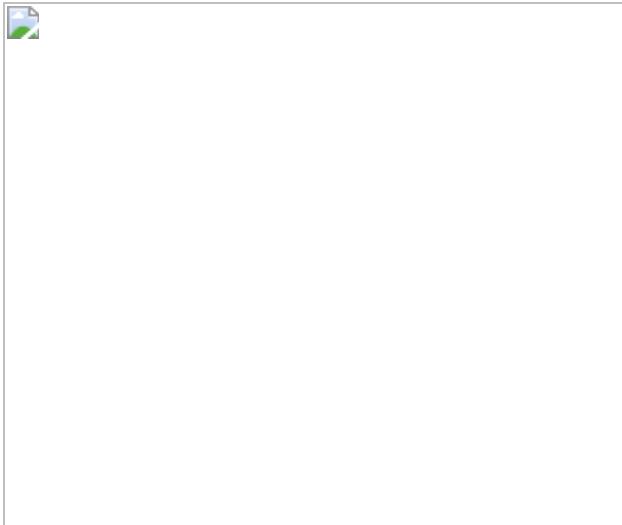
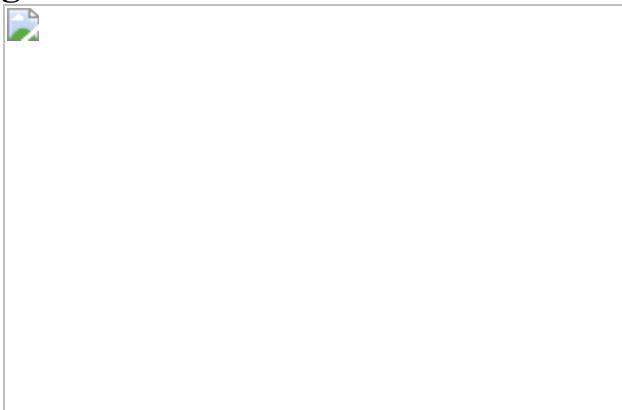


Fig. 3: Visualization of the bipolar outflow model and star-forming regions.



Data availability

The spectroscopic data analysed in this study are available from the Mikulski Archive for Space Telescopes (MAST) at
<https://archive.stsci.edu/>.

References

1. 1.

Manzano-King, C. M., Canalizo, G. & Sales, L. V. AGN-driven outflows in dwarf galaxies. *Astrophys. J.* **884**, 54 (2019).

2. 2.

Fabian, A. C. Observational evidence of active galactic nuclei feedback. *Annu. Rev. Astron. Astrophys.* **50**, 455–489 (2012).

3. 3.

Gaibler, V. et al. Jet-induced star formation in gas-rich galaxies. *Mon. Not. R. Astron. Soc.* **425**, 438–449 (2012).

4. 4.

Maiolino, R. et al. Star formation inside a galactic outflow. *Nature* **544**, 202–206 (2017).

5. 5.

Gallagher, R. et al. Widespread star formation inside galactic outflows. *Mon. Not. R. Astron. Soc.* **485**, 3409–3429 (2019).

6. 6.

Reines, A. E. et al. An actively accreting massive black hole in the dwarf starburst galaxy Henize 2-10. *Nature* **470**, 66–68 (2011).

7. 7.

Reines, A. E. & Adam, T. D. Parsec-scale radio emission from the low-luminosity active galactic nucleus in the dwarf starburst galaxy Henize 2-10. *Astrophys. J. Lett.* **750**, L24 (2012).

8. 8.

Reines, A. E. et al. Deep Chandra observations of the compact starburst galaxy Henize 2-10: X-rays from the massive black hole. *Astrophys. J. Lett.* **830**, L35 (2016).

9. 9.

Riffel, R. A. Evidence for an accreting massive black hole in He 2-10 from adaptive optics integral field spectroscopy. *Mon. Not. R. Astron. Soc.* **494**, 2004–2011 (2020).

10. 10.

Hebbar, P. R. et al. X-ray spectroscopy of the candidate AGNs in Henize 2-10 and NGC 4178: likely supernova remnants. *Mon. Not. R. Astron. Soc.* **485**, 5604–5615 (2019).

11. 11.

Cresci, G. et al. The MUSE view of He 2-10: No AGN ionization but a sparkling starburst. *Astron. Astrophys.* **604**, A101 (2017).

12. 12.

Kobulnicky, H. A. et al. Aperture synthesis observations of molecular and atomic gas in the Wolf-Rayet starburst galaxy. *Astron. J.* **110**, 116 (1995).

13. 13.

Mathewson, D. S. et al. A new oxygen-rich supernova remnant in the Large Magellanic Cloud. *Astrophys. J.* **242**, L73–L76 (1980).

14. 14.

Borkowski, K. J. et al. Asymmetric expansion of the youngest galactic supernova remnant G1. 9+ 0.3. *Astrophys. J. Lett.* **837**, L7 (2017).

15. 15.

Gower, A. C. & Hutchings, J. B. A precessing relativistic jet model for 3C 449. *Astrophys. J.* **258**, L63–L66 (1982).

16. 16.

Dunn, R. J. H., Fabian, A. C. & Sanders, J. S. Precession of the supermassive black hole in NGC 1275 (3C 84)? *Mon. Not. R. Astron. Soc.* **366**, 758–766 (2006).

17. 17.

Pringle, J. E. Self-induced warping of accretion discs. *Mon. Not. R. Astron. Soc.* **281**, 357–361 (1996).

18. 18.

Nixon, C. & King, A. Do jets precess... or even move at all? *Astrophys. J. Lett.* **765**, L7 (2013).

19. 19.

Kharb, P. et al. Double-peaked emission lines due to a radio outflow in KISSR 1219. *Astrophys. J.* **846**, 12 (2017).

20. 20.

Beck, S. C., Jean, L. T. & Michelle Consiglio, S. Dense molecular filaments feeding a starburst: ALMA maps of CO (3–2) in Henize 2-10. *Astrophys. J.* **867**, 165 (2018).

21. 21.

Lee, M. G. et al. Optical spectroscopy of supernova remnants in M81 and M82. *Astrophys. J.* **804**, 63 (2015).

22. 22.

Trump, J. R. et al. Accretion rate and the physical nature of unobscured active galaxies. *Astrophys. J.* **733**, 60 (2011).

23. 23.

Reines, A. E. et al. A new sample of (wandering) massive black holes in dwarf galaxies from high-resolution radio observations. *Astrophys.*

J. **888**, 36 (2020).

24. 24.

Molina, M. et al. Outflows, shocks, and coronal line emission in a radio-selected AGN in a dwarf galaxy. *Astrophys. J.* **910**, 5 (2021).

25. 25.

Allen, M. G. et al. The MAPPINGS III library of fast radiative shock models. *Astrophys. J. Suppl. Ser.* **178**, 20 (2008).

26. 26.

Silk, J. & Norman, C. Global star formation revisited. *Astrophys. J.* **700**, 262 (2009).

27. 27.

Silk, J. Unleashing positive feedback: linking the rates of star formation, supermassive black hole accretion, and outflows in distant galaxies. *Astrophys. J.* **772**, 112 (2013).

28. 28.

Penny, S. J. et al. SDSS-IV MaNGA: evidence of the importance of AGN feedback in low-mass galaxies. *Mon. Not. R. Astron. Soc.* **476**, 979–998 (2018).

29. 29.

Trump, J. R. et al. Spectropolarimetric evidence for radiatively inefficient accretion in an optically dull active galaxy. *Astrophys. J.* **732**, 23 (2011).

30. 30.

Santoro, F. et al. AGN-driven outflows and the AGN feedback efficiency in young radio galaxies. *Astron. Astrophys.* **644**, A54

(2020).

31. 31.

Constantin, A. et al. Probing the balance of AGN and star-forming activity in the local universe with ChaMP. *Astrophys. J.* **705**, 1336 (2009).

32. 32.

Baganoff, F. K. et al. Chandra X-ray spectroscopic imaging of Sagittarius A* and the central parsec of the galaxy. *Astrophys. J.* **591**, 891 (2003).

33. 33.

Nguyen, D. D. et al. Extended structure and fate of the nucleus in Henize 2-10. *Astrophys. J.* **794**, 34 (2014).

34. 34.

Greene, J. E., Strader, J. & Ho, L. C. Intermediate-mass black holes. *Annu. Rev. Astron. Astrophys.* **58**, 257–312 (2020).

35. 35.

Leitherer, C. et al. Starburst99: synthesis models for galaxies with active star formation. *Astrophys. J. Suppl. Ser.* **123**, 3 (1999).

36. 36.

Newville, M., Stensitzki, T., Allen, D. B. & Ingargiola, A. LMFIT: non-linear least-square minimization and curve-fitting for Python version 0.8.0. *Zenodo* <https://doi.org/10.5281/zenodo.11813> (2014).

37. 37.

Osterbrock, D. E., and Ferland, G. J. *Astrophysics of Gas Nebulae and Active Galactic Nuclei* (University Science Books, 2006).

38. 38.

Baldwin, J. A., Phillips, M. M. & Terlevich, R. Classification parameters for the emission-line spectra of extragalactic objects. *Publ. Astron. Soc. Pac.* **93**, 5 (1981).

39. 39.

Veilleux, S. & Osterbrock, D. E. Spectral classification of emission-line galaxies. *Astrophys. J. Suppl. Ser.* **63**, 295–310 (1987).

40. 40.

Kewley, L. J. et al. The host galaxies and classification of active galactic nuclei. *Mon. Not. R. Astron. Soc.* **372**, 961–976 (2006).

41. 41.

Kauffmann, G. et al. The host galaxies of active galactic nuclei. *Mon. Not. R. Astron. Soc.* **346**, 1055–1077 (2003).

42. 42.

Kewley, L. J. et al. Theoretical modeling of starburst galaxies. *Astrophys. J.* **556**, 121 (2001).

43. 43.

Martín-Hernández, N. L. et al. High spatial resolution mid-infrared spectroscopy of the starburst galaxies NGC 3256, II Zw 40 and Henize 2-10. *Astron. Astrophys.* **455**, 853–870 (2006).

44. 44.

Chandar, R. et al. The stellar content of Henize 2-10 from space telescope imaging spectrograph ultraviolet spectroscopy. *Astrophys. J.* **586**, 939 (2003).

45. 45.

Nawaz, M. A. et al. Jet–intracluster medium interaction in Hydra A–II. The effect of jet precession. *Mon. Not. R. Astron. Soc.* **458**, 802–815 (2016).

46. 46.

Cielo, S. et al. Feedback from reorienting AGN jets-I. Jet–ICM coupling, cavity properties and global energetics. *Astron. Astrophys.* **617**, A58 (2018).

Acknowledgements

We are grateful to M. Molina for useful discussions regarding shocks. We also thank M. Whittle and K. Johnson for their assistance with the HST/STIS proposal while A.E.R. was a graduate student at the University of Virginia, and for subsequent discussions. Support for Program number HST-GO-12584.006-A was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. A.E.R. also acknowledges support for this work provided by NASA through EPSCoR grant number 80NSSC20M0231. Z.S. acknowledges support for this project from the Montana Space Grant Consortium.

Author information

Affiliations

1. eXtreme Gravity Institute, Department of Physics, Montana State University, Bozeman, MT, USA

Zachary Schutte & Amy E. Reines

Contributions

Z.S. reduced and analysed the STIS data and compared the results with models. A.E.R. led the HST/STIS proposal and helped with the data reduction. Both authors worked on the interpretation of the results and the writing of the paper.

Corresponding author

Correspondence to [Zachary Schutte](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

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

Extended data figures and tables

[Extended Data Fig. 1 Raw 2D spectra showing the \[OI\]6300 emission line at the location of the nucleus in the EW slit orientation.](#)

The location of the nucleus is indicated by white circles and the two images correspond to the two dithered sub-exposures.

[Extended Data Fig. 2 Combined 2D spectra showing the \[OI\]6300 emission line at the location of the nucleus in the EW slit orientation.](#)

Same as Extended Data Fig. 1 but showing the reduced 2D image with the dithered sub-exposures combined.

Extended Data Fig. 3 The electron density, n_e , along the EW slit orientation.

We measure the electron density along the EW slit from the ratio of [SII]6716/[SII]6731 and find the electron density ranges from $\sim 10^{2.5} - 10^{4} \text{ cm}^{-3}$, which is within the range the [SII] ratio is sensitive to density. The high densities are consistent with those predicted by optical emission line diagnostics derived from the Allen et al.²⁵ shock models.

Extended Data Fig. 4 The spatial extraction regions taken along the EW slit orientation.

We place these regions on optical emission line diagnostic diagrams (Extended Data Figs. 5–7). Top panel: the extraction regions are shown on the narrow band H α + continuum image from HST to highlight the ionized gas features that several of the spatial extractions probe. Bottom panel: the extraction regions are shown on the archival 0.8 micron HST image, showing young star clusters that the EW slit orientation passes through.

Extended Data Fig. 5 Narrow emission line diagnostic diagrams showing various extraction regions along the EW slit orientation (see Extended Data Fig. 4).

The nucleus (yellow point) falls in the Seyfert region of the [OI]/H α) diagram. The young star-forming region ~70 pc to the east of the low-luminosity AGN is depicted with a blue triangle and star for the primary emission line component and the blue-shifted secondary component, respectively. [OI] is not detected in all of the regions.

Extended Data Fig. 6 Optical emission line diagnostics from the shock and shock+precursor models with varying gas density.

We place the spatial extractions from the EW slit orientation shown in Extended Data Fig. 4 on a grid of shock excitation models (presented in Allen et al.²⁵ with varying gas density ($n = 0.01\text{-}1000 \text{ cm}^{-3}$) and shock velocity ($v = 100\text{-}600 \text{ km/s}$). We fix the transverse magnetic field to be $b = 1 \mu \text{G}$ and assume solar metallicity.

Extended Data Fig. 7 Optical emission line diagnostics from the shock and shock+precursor models with varying magnetic field.

The models (presented in Allen et al.²⁵) are shown as a grid with dashed blue lines indicating constant shock velocity and dashed black lines indicating constant transverse magnetic field. For these models, the density is fixed to $n = 1000 \text{ cm}^{-3}$ and the transverse magnetic field parameter is allowed to vary from $b = 0.01\text{-}32 \mu \text{G}$.

Extended Data Fig. 8 A diagram of the toy model of the bipolar outflow generated by the low-luminosity AGN in Henize 2-10.

Our simple model depends on the outflow velocity of the ionized gas (v_{outflow}), the angle the outflow makes with its precession axis (θ) and the angular frequency with which the outflow precesses (ω). Similar models have been used to describe the bending seen in large radio jets^{15,16}.

Extended Data Table 1 Summary of observational results regarding the nature of the nucleus in Henize 2-10

Supplementary information

Peer Review File

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Schutte, Z., Reines, A.E. Black-hole-triggered star formation in the dwarf galaxy Henize 2-10. *Nature* **601**, 329–333 (2022).
<https://doi.org/10.1038/s41586-021-04215-6>

- Received: 29 June 2021
- Accepted: 08 November 2021
- Published: 19 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04215-6>

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

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

- Article
- [Published: 12 January 2022](#)

Star formation near the Sun is driven by expansion of the Local Bubble

- [Catherine Zucker](#) [ORCID: orcid.org/0000-0002-2250-730X^{1,2}](#),
- [Alyssa A. Goodman](#) [ORCID: orcid.org/0000-0003-1312-0477¹](#),
- [João Alves](#) [ORCID: orcid.org/0000-0002-4355-0921³](#),
- [Shmuel Bialy](#) [ORCID: orcid.org/0000-0002-0404-003X^{1,4}](#),
- [Michael Foley](#) [ORCID: orcid.org/0000-0002-6747-2745¹](#),
- [Joshua S. Speagle](#) [ORCID: orcid.org/0000-0003-2573-9832^{5,6,7}](#),
- [Josefa Großschedl](#) [ORCID: orcid.org/0000-0002-6895-2804³](#),
- [Douglas P. Finkbeiner^{1,8}](#),
- [Andreas Burkert^{9,10}](#),
- [Diana Khimey¹](#) &
- [Cameren Swiggum](#) [ORCID: orcid.org/0000-0001-9201-5995^{3,11}](#)

[Nature](#) volume 601, pages 334–337 (2022)

- 5741 Accesses
- 1030 Altmetric
- [Metrics details](#)

Subjects

- [Interstellar medium](#)
- [Stars](#)

Abstract

For decades we have known that the Sun lies within the Local Bubble, a cavity of low-density, high-temperature plasma surrounded by a shell of cold, neutral gas and dust^{1,2,3}. However, the precise shape and extent of this shell^{4,5}, the impetus and timescale for its formation^{6,7}, and its relationship to nearby star formation⁸ have remained uncertain, largely due to low-resolution models of the local interstellar medium. Here we report an analysis of the three-dimensional positions, shapes and motions of dense gas and young stars within 200 pc of the Sun, using new spatial^{9,10,11} and dynamical constraints¹². We find that nearly all of the star-forming complexes in the solar vicinity lie on the surface of the Local Bubble and that their young stars show outward expansion mainly perpendicular to the bubble's surface. Tracebacks of these young stars' motions support a picture in which the origin of the Local Bubble was a burst of stellar birth and then death (supernovae) taking place near the bubble's centre beginning approximately 14 Myr ago. The expansion of the Local Bubble created by the supernovae swept up the ambient interstellar medium into an extended shell that has now fragmented and collapsed into the most prominent nearby molecular clouds, in turn providing robust observational support for the theory of supernova-driven star formation.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

[Subscribe to Journal](#)

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

[Buy article](#)

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: A 3D spatial view of the solar neighbourhood.

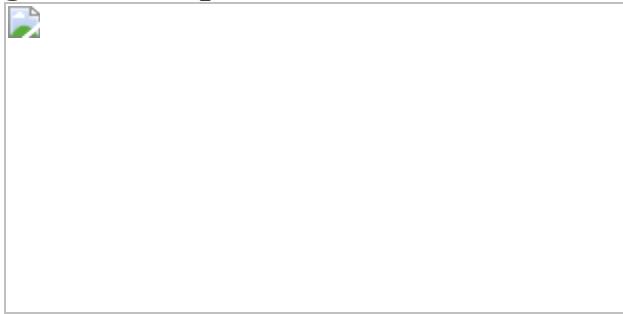
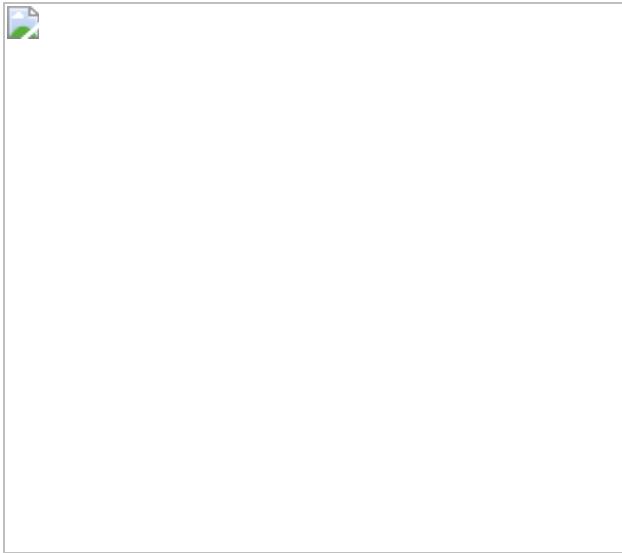


Fig. 2: The evolution of the Local Bubble and sequential star formation at the surface of its expanding shell.



Data availability

The datasets generated and/or analysed during the current study are publicly available on the Harvard Dataverse (https://dataverse.harvard.edu/dataverse/local_bubble_star_formation/), including Extended Data Table 1 (<https://doi.org/10.7910/DVN/ZU97QD>), Extended Data Table 2 (<https://doi.org/10.7910/DVN/1VT8BC>), per-star data for individual stellar cluster members (<https://doi.org/10.7910/DVN/1UPMDX>) and the cluster tracebacks (<https://doi.org/10.7910/DVN/E8PQOD>).

Code availability

The results generated in this work are based on publicly available software packages and do not involve the extensive use of custom code. Given each star's reported *Gaia* data, we use the astropy³⁸ package to obtain the Heliocentric Galactic Cartesian positions and velocities. The extreme deconvolution algorithm in the astroML⁵¹ package is used to estimate the mean 3D positions and velocities of the stellar clusters. The Orbit functionality in the galpy⁴⁰ package is used to perform the dynamical tracebacks. The dynesty⁴³ package is used to fit the analytic superbubble expansion model and determine the best-fit parameters governing the Local Bubble's evolution.

References

1. 1.

Cox, D. P. & Reynolds, R. J. The local interstellar medium. *Annu. Rev. Astron. Astrophys.* **25**, 303–344 (1987).

2. 2.

Lucke, P. B. The distribution of color excesses and interstellar reddening material in the solar neighborhood. *Astron. Astrophys.* **64**, 367–377 (1978).

3. 3.

Sanders, W. T., Kraushaar, W. L., Nousek, J. A. & Fried, P. M. Soft diffuse X-rays in the southern galactic hemisphere. *Astrophys. J. Lett.* **217**, L87–L91 (1977).

4. 4.

Lallement, R., Welsh, B. Y., Vergely, J. L., Crifo, F. & Sfeir, D. 3D mapping of the dense interstellar gas around the Local Bubble. *Astron. Astrophys.* **411**, 447–464 (2003).

5. 5.

Welsh, B. Y., Lallement, R., Vergely, J.-L. & Raimond, S. New 3D gas density maps of NaI and CaII interstellar absorption within 300 pc. *Astron. Astrophys.* **510**, A54 (2010).

6. 6.

Fuchs, B., Breitschwerdt, D., de Avillez, M. A., Dettbarn, C. & Flynn, C. The search for the origin of the Local Bubble redivivus. *Mon. Not. R. Astron. Soc.* **373**, 993–1003 (2006).

7. 7.

Breitschwerdt, D. et al. The locations of recent supernovae near the Sun from modelling ^{60}Fe transport. *Nature*. **532**, 73–76 (2016).

8. 8.

Frisch, P. & Dwarkadas, V. V. in *Handbook of Supernovae* (eds Alsabti, A. W. & Murdin, P.) 2253–2285 (Springer International Publishing, 2017).

9. 9.

Leike, R. H., Glatzle, M. & Enßlin, T. A. Resolving nearby dust clouds. *Astron. Astrophys.* **639**, A138 (2020).

10. 10.

Lallement, R. et al. Gaia-2MASS 3D maps of Galactic interstellar dust within 3 kpc. *Astron. Astrophys.* **625**, A135 (2019).

11. 11.

Zucker, C. et al. On the three-dimensional structure of local molecular clouds. *Astrophys. J.* **919**, 35 (2021).

12. 12.

Lindegren, L. et al. Gaia Early Data Release 3 – the astrometric solution. *Astron. Astrophys. Suppl. Ser.* **649**, A2 (2021).

13. 13.

Pelgrims, V., Ferrière, K., Boulanger, F., Lallement, R. & Montier, L. Modeling the magnetized Local Bubble from dust data. *Astron. Astrophys.* **636**, A17 (2020).

14. 14.

Welsh, B. Y., Sfeir, D. M., Sirk, M. M. & Lallement, R. EUV mapping of the local interstellar medium: the Local Chimney revealed? *Astron.*

Astrophys. **352**, 308–316 (1999).

15. 15.

Bialy, S. et al. The Per-Tau Shell: a giant star-forming spherical shell revealed by 3D dust observations. *Astrophys. J. Lett.* **919**, L5 (2021).

16. 16.

Alves, J. et al. A Galactic-scale gas wave in the solar neighbourhood. *Nature*. **578**, 237–239 (2020).

17. 17.

Großschedl, J. E., Alves, J., Meingast, S. & Herbst-Kiss, G. 3D dynamics of the Orion cloud complex – discovery of coherent radial gas motions at the 100-pc scale. *Astron. Astrophys. Suppl. Ser.* **647**, A91 (2021).

18. 18.

Perrot, C. A. & Grenier, I. A. 3D dynamical evolution of the interstellar gas in the Gould Belt. *Astron. Astrophys. Suppl. Ser.* **404**, 519–531 (2003).

19. 19.

Dzib, S. A., Loinard, L., Ortiz-León, G. N., Rodríguez, L. F. & Galli, P. A. B. Distances and kinematics of Gould Belt star-forming regions with Gaia DR2 results. *Astrophys. J.* **867**, 151 (2018).

20. 20.

Kerr, R. M. P., Rizzuto, A. C., Kraus, A. L. & Offner, S. S. R. Stars with Photometrically Young Gaia Luminosities Around the Solar System (SPYGLASS). I. Mapping young stellar structures and their star formation histories. *Astrophys. J.* **917**, 23 (2021).

21. 21.

Maíz-Apellániz, J. The origin of the Local Bubble. *Astrophys. J. Lett.* **560**, L83–L86 (2001).

22. 22.

El-Badry, K., Ostriker, E. C., Kim, C.-G., Quataert, E. & Weisz, D. R. Evolution of supernovae-driven superbubbles with conduction and cooling. *Mon. Not. R. Astron. Soc.* **490**, 1961–1990 (2019).

23. 23.

Inutsuka, S.-I., Inoue, T., Iwasaki, K. & Hosokawa, T. The formation and destruction of molecular clouds and galactic star formation. An origin for the cloud mass function and star formation efficiency. *Astron. Astrophys.* **580**, A49 (2015).

24. 24.

Dawson, J. R. The supershell–molecular cloud connection: large-scale stellar feedback and the formation of the molecular ISM. *Publ. Astron. Soc. Aust.* **30**, e025 (2013).

25. 25.

Cox, D. P. & Smith, B. W. Large-scale effects of supernova remnants on the Galaxy: generation and maintenance of a hot network of tunnels. *Astrophys. J. Lett.* **189**, L105–L108 (1974).

26. 26.

McKee, C. F. & Ostriker, J. P. A theory of the interstellar medium: three components regulated by supernova explosions in an inhomogeneous substrate. *Astrophys. J.* **218**, 148–169 (1977).

27. 27.

Kim, C.-G., Ostriker, E. C. & Raileanu, R. Superbubbles in the multiphase ISM and the loading of Galactic winds. *Astrophys. J.* **834**, 25 (2017).

28. 28.

Galli, P. A. B. et al. Lupus DANCe. Census of stars and 6D structure with Gaia-DR2 data. *Astron. Astrophys.* **643**, A148 (2020).

29. 29.

Grasser, N. et al. The ρ Oph region revisited with Gaia EDR3. *Astron. Astrophys.* **652**, A2 (2021)

30. 30.

Galli, P. A. B. et al. Chamaeleon DANCe. Revisiting the stellar populations of Chamaeleon I and Chamaeleon II with Gaia-DR2 data. *Astron. Astrophys.* **646**, A46 (2021).

31. 31.

Galli, P. A. B. et al. Corona-Australis DANCe. I. Revisiting the census of stars with Gaia-DR2 data. *Astron. Astrophys.* **634**, A98 (2020).

32. 32.

Krolikowski, D. M., Kraus, A. L. & Rizzuto, A. C. Gaia EDR3 reveals the substructure and complicated star formation history of the Greater Taurus-Auriga star-forming complex. *Astron. J.* **162**, 3 (2021).

33. 33.

Gagné, J. & Faherty, J. K. BANYAN. XIII. A first look at nearby young associations with Gaia Data Release 2. *Astrophys. J.* **862**, 138 (2018).

34. 34.

Gagné, J. et al. BANYAN. XI. The BANYAN Σ multivariate Bayesian algorithm to identify members of young associations with 150 pc. *Astrophys. J.* **856**, 23 (2018).

35. 35.

Ortiz-León, G. N. et al. The Gould’s Belt Distances Survey (GOBELINS). V. Distances and kinematics of the Perseus Molecular Cloud. *Astrophys. J.* **865**, 73 (2018).

36. 36.

Herczeg, G. J. et al. An initial overview of the extent and structure of recent star formation within the Serpens molecular cloud using Gaia Data Release 2. *Astrophys. J.* **878**, 111 (2019).

37. 37.

Fabricius, C. et al. Gaia Early Data Release 3 – catalogue validation. *Astron. Astrophys. Suppl. Ser.* **649**, A5 (2021).

38. 38.

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

39. 39.

Bovy, J., Hogg, D. W. & Roweis, S. T. Extreme deconvolution: inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *Ann. Appl. Stat.* **5**, 1657–1677 (2011).

40. 40.

Bovy, J. galpy: a Python library for Galactic dynamics. *Astrophys. J. Supp.* **216**, 29 (2015).

41. 41.

Kerr, F. J. & Lynden-Bell, D. Review of galactic constants. *Mon. Not. R. Astron. Soc.* **221**, 1023–1038 (1986).

42. 42.

Kamdar, H., Conroy, C. & Ting, Y.-S. Stellar streams in the Galactic disk: predicted lifetimes and their utility in measuring the galactic potential. Preprint at <https://arxiv.org/abs/2106.02050v1> (2021).

43. 43.

Speagle, J. S. dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.* **493**, 3132–3158 (2020).

44. 44.

Salpeter, E. E. The luminosity function and stellar evolution. *Astrophys. J.* **121**, 161 (1955).

45. 45.

Gontcharov, G. & Mosenkov, A. Interstellar polarization and extinction in the Local Bubble and the Gould Belt. *Mon. Not. R. Astron. Soc.* **483**, 299–314 (2019).

46. 46.

Dehnen, W. & Binney, J. J. Local stellar kinematics from Hipparcos data. *Mon. Not. R. Astron. Soc.* **298**, 387–394 (1998).

47. 47.

Francis, C. & Anderson, E. Calculation of the local standard of rest from 20574 local stars in the New Hipparcos Reduction with known radial velocities. *New Astron.* **14**, 615–629 (2009).

48. 48.

Wang, F. et al. Local stellar kinematics and Oort constants from the LAMOST A-type stars. *Mon. Not. R. Astron. Soc.* **504**, 199–207 (2021).

49. 49.

Reid, M. J. et al. Trigonometric parallaxes of high-mass star-forming regions: our view of the Milky Way. *Astrophys. J.* **885**, 131 (2019).

50. 50.

Schönrich, R., Binney, J. & Dehnen, W. Local kinematics and the local standard of rest. *Mon. Not. R. Astron. Soc.* **403**, 1829–1833 (2010).

51. 51.

VanderPlas, J., Connolly, A. J., Ivezić, Ž. & Gray, A. Introduction to astroML: machine learning for astrophysics. In *Proc. 2012 Conference on Intelligent Data Understanding* 47–54 (IEEE, 2012).

Acknowledgements

The visualization, exploration and interpretation of data presented in this work were made possible using the glue visualization software, supported under NSF grant numbers OAC-1739657 and CDS&E:AAG-1908419. The interactive figures were made possible by the plot.ly python library. D.P.F. acknowledges support by NSF grant AST-1614941 ‘Exploring the Galaxy: 3-Dimensional Structure and Stellar Streams’. D.P.F., A.A.G. and C.Z. acknowledge support by NASA ADAP grant 80NSSC21K0634 ‘Knitting Together the Milky Way: An Integrated Model of the Galaxy’s Stars, Gas, and Dust’. A.B. acknowledges support by the Excellence Cluster ORIGINS, which is funded by the German Research Foundation (DFG) under Germany’s Excellence Strategy -EXC-2094-390783311. J.A. acknowledges support from the Data Science Research Centre and the TURIS Research Platform of the University of Vienna. J.G. acknowledges funding by the Austrian Research Promotion Agency (FFG) under project number 873708. C.Z. acknowledges that support for this work was provided by NASA through the NASA Hubble Fellowship grant number HST-HF2-51498.001 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. C.Z., A.A.G., J.A. and S.B.

acknowledge Interstellar Institute's program 'The Grand Cascade' and the Paris-Saclay University's Institut Pascal for hosting discussions that encouraged the development of the ideas behind this work.

Author information

Affiliations

1. Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA

Catherine Zucker, Alyssa A. Goodman, Shmuel Bialy, Michael Foley, Douglas P. Finkbeiner & Diana Khimey

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

Catherine Zucker

3. Department of Astrophysics, University of Vienna, Wien, Austria

João Alves, Josefa Großschedl & Cameren Swiggum

4. Department of Astronomy, University of Maryland, College Park, MD, USA

Shmuel Bialy

5. Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

Joshua S. Speagle

6. David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, Toronto, Ontario, Canada

Joshua S. Speagle

7. Dunlap Institute for Astronomy & Astrophysics, University of
Toronto, Toronto, Ontario, Canada

Joshua S. Speagle

8. Harvard University Department of Physics, Cambridge, MA, USA

Douglas P. Finkbeiner

9. University Observatory Munich, Munich, Germany

Andreas Burkert

10. Max-Planck-Institut für extraterrestrische Physik, Garching, Germany

Andreas Burkert

11. Department of Astronomy, University of Wisconsin, Madison, WI,
USA

Cameren Swiggum

Contributions

C.Z. led the work and wrote the majority of the text. All authors contributed to the text. C.Z., A.A.G. and J.A. led interpretation of the observational results, aided by S.B., M.F. and A.B. who helped interpret their significance in light of theoretical models for supernova-driven star formation. C.Z. and A.A.G. led the visualization efforts. J.S.S. and D.P.F. helped shape the statistical modelling of the Local Bubble's expansion. C.Z., A.A.G. and J.S.S. contributed to the software used in this work. J.G. and C.S. provided data for and the subsequent interpretation of the 3D kinematics of the Orion region. D.K. helped to develop the code used to model the 3D positions and motions of stellar clusters described in the Methods.

Corresponding author

Correspondence to [Catherine Zucker](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Joanne Dawson 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 1D and 2D marginal distributions \(“corner plot”\) of the model parameters governing the evolution of the Local Bubble’s expanding shell.](#)

Parameters include the time since the first explosion (i.e. the age of the Local Bubble), t_{exp} , the density of the ambient medium the bubble is expanding into, n_0 , the time between supernova explosions powering its growth, t_{SNe} , and the thickness/uncertainty on the expanding shell radius R . In the 1D distributions, the vertical dashed lines denote the median and 1 σ errors, while in the 2D distributions, we show the 0.5 σ , 1 σ , 1.5 σ , and 2 σ contours.

Extended Data Fig. 2 Temporal evolution of the Local Bubble, based on the fit to the dynamical tracebacks and the analytic expansion model²² summarized in the Methods section.

Panel a) The evolution of the Local Bubble's expansion velocity v_{exp} .

Panel b) The evolution of the Local Bubble's shell radius R_{shell} . **Panel c)** The evolution of the average momentum injection per supernova \hat{p} . The thick purple line represents the median fit, while the thin purple lines represent random samples. We estimate a current radius of $(165 \pm 6) \text{ pc}$ and current expansion velocity of $(6.7 \pm 0.4)^{+0.5} \text{ km/s}$, corresponding to time $t=0$ Myr (the present day).

Extended Data Fig. 3 PDF of the estimate of the number of supernovae required to power the Local Bubble's expansion.

The estimate is obtained by comparing the shell's present-day momentum to the average momentum injected by supernovae.

Extended Data Fig. 4 Analysis of the stellar tracebacks of the UCL and LCC clusters, whose progenitors were likely responsible for the supernovae that created the Local Bubble.

The scatter points indicate the positions of the current cluster members of UCL and LCC, which are colored as a function of time (spanning the present day in pink to 30 Myr ago in black). **Panel a:** Using Hipparcos data and adopting a solar peculiar motion $(U_{\odot}, V_{\odot}, W_{\odot}) = (10.0, 5.2, 7.2) \text{ km/s}^{46}$, previous literature^{6,7} find that UCL and LCC were born outside the Local Bubble (black trace⁴) 15 Myr ago and only entered its present-day boundary in the past 5 Myr (reproduced from Fig. 6 in ref. ⁶). **Panel b:** We attempt to reproduce the results from previous literature^{6,7} using the same data and solar motion, but are unable to account for the curvature of the tracebacks, finding the UCL and LCC formed just inside its northern boundary 15 Myr ago. **Panel c:** Using a different value for the solar motion, $(U_{\odot}, V_{\odot}, W_{\odot}) = (10.0, 15.4, 7.8) \text{ km/s}^{41}$ but the same Hipparcos data, we find that UCL and LCC were born near the center of the Local Bubble.

Panel d: Finally, using updated *Gaia* data but the same adopted solar motion used in panel c. (U_\odot , V_\odot , W_\odot) = (10.0, 15.4, 7.8) km/s⁴¹, we also find that UCL and LCC were born near the center of the bubble, given an updated model for its surface¹³.

Extended Data Table 1 Summary of the 3D positions and 3D velocities of young stellar clusters within 400 pc of the Sun

Extended Data Table 2 Temporal evolution of cluster births at the surface of the Local Bubble's expanding shell

Supplementary information

Peer Review File

Supplementary Figure 1

Interactive 3D visualization of dense gas and young stars on the Local Bubble's surface. This figure is the interactive 3D counterpart to Fig. 1. The figure supports interactive panning, zooming and rotation. Individual data layers can be toggled on/off by clicking on the layer in the legend on the right-hand side of the figure. The surface of the Local Bubble¹³ is shown in purple. The short squiggly coloured lines (or ‘skeletons’) demarcate the 3D spatial morphology of dense gas in prominent nearby molecular clouds¹¹. The 3D cones indicate the positions of young stellar clusters, with the apex of the cone pointing in the direction of stellar motion. The Sun is marked with a yellow cross. We also overlay the morphology of the 3D dust (grey blobby shapes⁹) and the models for two Galactic scale features—the Radcliffe Wave (red)¹⁶ and the Split (blue)¹⁰. The Per-Tau Superbubble¹⁵ (green sphere) is also overlaid.

Supplementary Figure 2

Interactive 3D visualization of the Local Bubble’s expansion. This figure is the interactive 3D counterpart to Fig. 2. The figure supports interactive panning, zooming and rotation. Individual data layers can be toggled on/off by clicking on the layer in the legend on the right-hand side of the figure.

Stellar cluster tracebacks are shown with the coloured paths. Before the cluster birth, the tracebacks are shown as semi-transparent circles meant to guide the eye, since our modelling is insensitive to the dynamics of the gas before its conversion into stars. After the cluster birth, the tracebacks are shown with filled circles and terminate in a large dot, which marks the cluster's current position. For time snapshots \leq 14 Myr ago, we overlay a model for the evolution of the Local Bubble (purple sphere), as derived in the Methods. Click 'Play Forward' to see the Local Bubble evolve starting 17 Myr ago and progressing forwards to the present day. Click 'Play Backward' to see the evolution in reverse. Click 'Pause' to stop the animation. Alternatively, drag the time slider back and forth to view the Local Bubble's expansion at any time. To jump to epochs of particular interest, click on any of the 'action' buttons (for example, 'UCL Born') on the right-hand side of the figure.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Zucker, C., Goodman, A.A., Alves, J. *et al.* Star formation near the Sun is driven by expansion of the Local Bubble. *Nature* **601**, 334–337 (2022).
<https://doi.org/10.1038/s41586-021-04286-5>

- Received: 18 August 2021
- Accepted: 26 November 2021
- Published: 12 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04286-5>

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/s41586-021-04286-5>

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

- Article
- [Published: 19 January 2022](#)

Fast universal quantum gate above the fault-tolerance threshold in silicon

- [Akito Noiri](#) [ORCID: orcid.org/0000-0001-9145-0303¹](#),
- [Kenta Takeda](#) [ORCID: orcid.org/0000-0003-1240-1103¹](#),
- [Takashi Nakajima](#) [ORCID: orcid.org/0000-0001-6759-6441¹](#),
- [Takashi Kobayashi](#) [ORCID: orcid.org/0000-0003-2841-8129²](#),
- [Amir Sammak^{3,4}](#),
- [Giordano Scappucci](#) [ORCID: orcid.org/0000-0003-2512-0079^{3,5}](#) &
- [Seigo Tarucha](#) [ORCID: orcid.org/0000-0001-7465-0135^{1,2}](#)

[Nature](#) volume **601**, pages 338–342 (2022)

- 2787 Accesses
- 1 Citations
- 223 Altmetric
- [Metrics details](#)

Subjects

- [Quantum dots](#)
- [Quantum information](#)
- [Qubits](#)

Abstract

Fault-tolerant quantum computers that can solve hard problems rely on quantum error correction¹. One of the most promising error correction codes is the surface code², which requires universal gate fidelities exceeding an error correction threshold of 99 per cent³. Among the many qubit platforms, only superconducting circuits⁴, trapped ions⁵ and nitrogen-vacancy centres in diamond⁶ have delivered this requirement. Electron spin qubits in silicon^{7,8,9,10,11,12,13,14,15} are particularly promising for a large-scale quantum computer owing to their nanofabrication capability, but the two-qubit gate fidelity has been limited to 98 per cent owing to the slow operation¹⁶. Here we demonstrate a two-qubit gate fidelity of 99.5 per cent, along with single-qubit gate fidelities of 99.8 per cent, in silicon spin qubits by fast electrical control using a micromagnet-induced gradient field and a tunable two-qubit coupling. We identify the qubit rotation speed and coupling strength where we robustly achieve high-fidelity gates. We realize Deutsch–Jozsa and Grover search algorithms with high success rates using our universal gate set. Our results demonstrate universal gate fidelity beyond the fault-tolerance threshold and may enable scalable silicon quantum computers.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Two-qubit system.

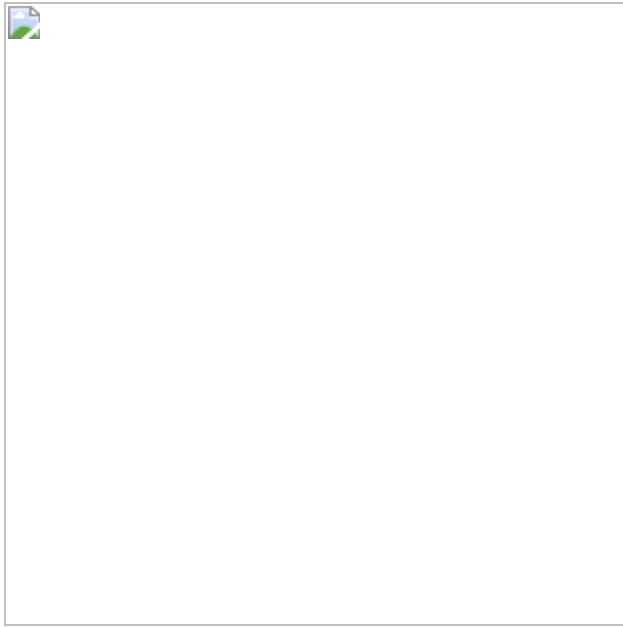


Fig. 2: Characterization of universal quantum control performances by randomized benchmarking.



Fig. 3: Rabi frequency dependence of single-qubit and two-qubit primitive gate fidelities.

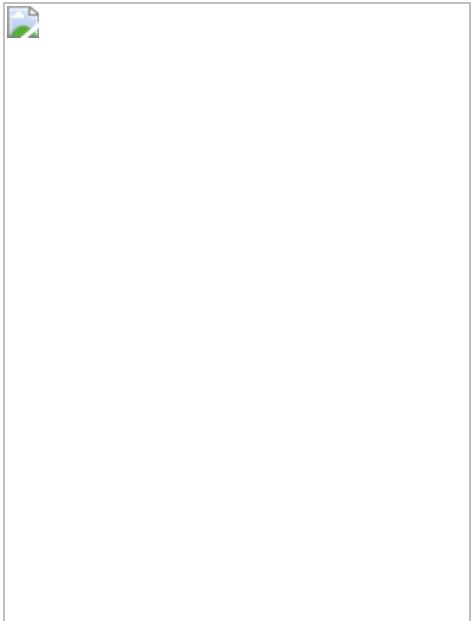
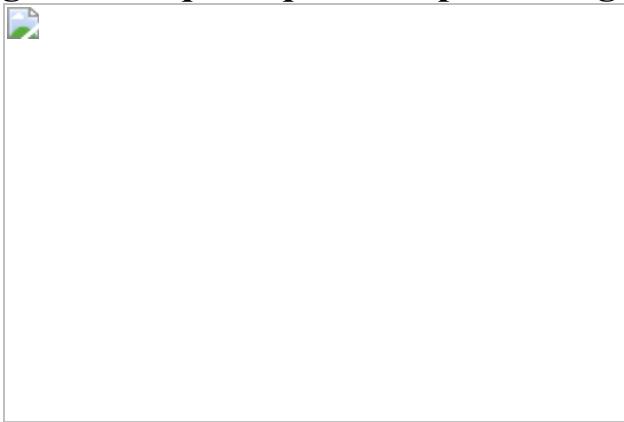


Fig. 4: Two-qubit quantum processing.



Data availability

All data in this study are available from the Zenodo repository at
<https://doi.org/10.5281/zenodo.5508362>.

References

1. 1.

Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).

2. 2.

Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).

3. 3.

Wang, D. S., Fowler, A. G. & Hollenberg, L. C. L. Surface code quantum computing with error rates over 1%. *Phys. Rev. A* **83**, 020302(R) (2011).

4. 4.

Barends, R. et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500–503 (2014).

5. 5.

Ballance, C. J., Harty, T. P., Linke, N. M., Sepiol, M. A. & Lucas, D. M. High-fidelity quantum logic gates using trapped-ion hyperfine qubits. *Phys. Rev. Lett.* **117**, 060504 (2016).

6. 6.

Rong, X. et al. Experimental fault-tolerant universal quantum gates with solid-state spins under ambient conditions. *Nat. Commun.* **6**, 8748 (2015).

7. 7.

Veldhorst, M. et al. An addressable quantum dot qubit with fault-tolerant control-fidelity. *Nat. Nanotechnol.* **9**, 981–985 (2014).

8. 8.

Yoneda, J. et al. A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%. *Nat. Nanotechnol.* **13**, 102–106 (2018).

9. 9.

Veldhorst, M. et al. A two-qubit logic gate in silicon. *Nature* **526**, 410–414 (2015).

10. 10.

Petit, L. et al. Universal quantum logic in hot silicon qubits. *Nature* **580**, 355–359 (2020).

11. 11.

Yang, C. H. et al. Operation of a silicon quantum processor unit cell above one kelvin. *Nature* **580**, 350–354 (2020).

12. 12.

Li, R. et al. A crossbar network for silicon quantum dot qubits. *Sci. Adv.* **4**, eaar3960 (2018).

13. 13.

Vandersypen, L. M. K. et al. Interfacing spin qubits in quantum dots and donors—hot, dense and coherent. *npj Quantum Inf.* **3**, 34 (2017).

14. 14.

Jones, C. et al. Logical qubit in a linear array of semiconductor quantum dots. *Phys. Rev. X* **8**, 21058 (2018).

15. 15.

Zajac, D. M., Hazard, T. M., Mi, X., Nielsen, E. & Petta, J. R. Scalable gate architecture for a one-dimensional array of semiconductor spin qubits. *Phys. Rev. Appl.* **6**, 054013 (2016).

16. 16.

Huang, W. et al. Fidelity benchmarks for two-qubit gates in silicon. *Nature* **569**, 532–536 (2019).

17. 17.

Yang, C. H. et al. Silicon qubit fidelities approaching incoherent noise limits via pulse engineering. *Nat. Electron.* **2**, 151–158 (2019).

18. 18.

Zajac, D. M. et al. Resonantly driven CNOT gate for electron spins. *Science* **359**, 439–442 (2018).

19. 19.

Watson, T. F. et al. A programmable two-qubit quantum processor in silicon. *Nature* **555**, 633–637 (2018).

20. 20.

Xue, X. et al. Benchmarking gate fidelities in Si/SiGe two-qubit device. *Phys. Rev. X* **9**, 021011 (2019).

21. 21.

Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120 (1998).

22. 22.

Petta, J. R. et al. Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005).

23. 23.

Reed, M. D. et al. Reduced sensitivity to charge noise in semiconductor spin qubits via symmetric operation. *Phys. Rev. Lett.* **116**, 110402 (2016).

24. 24.

Takeda, K., Noiri, A., Yoneda, J., Nakajima, T. & Tarucha, S. Resonantly driven singlet–triplet spin qubit in silicon. *Phys. Rev. Lett.* **124**, 117701 (2020).

25. 25.

Sigillito, A. J., Gullans, M. J., Edge, L. F., Borselli, M. & Petta, J. R. Coherent transfer of quantum information in silicon using resonant SWAP gates. *npj Quantum Inf.* **5**, 110 (2019).

26. 26.

Deutsch, D. & Jozsa, R. Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A* **439**, 553–558 (1992).

27. 27.

Grover, L. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79**, 325–328 (1997).

28. 28.

Yoneda, J. et al. Robust micro-magnet design for fast electrical manipulations of single spins in quantum dots. *Appl. Phys. Express* **8**, 084401 (2015).

29. 29.

Elzerman, J. M. et al. Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430**, 431–435 (2004).

30. 30.

Morello, A. et al. Single-shot readout of an electron spin in silicon. *Nature* **467**, 687–691 (2010).

31. 31.

Russ, M. et al. High-fidelity quantum gates in Si/SiGe double quantum dots. *Phys. Rev. B* **97**, 085421 (2018).

32. 32.

Nakajima, T. et al. Coherence of a driven electron spin qubit actively decoupled from quasi-static noise. *Phys. Rev. X* **10**, 11060 (2020).

33. 33.

Takeda, K. et al. A fault-tolerant addressable spin qubit in a natural silicon quantum dot. *Sci. Adv.* **2**, e1600694 (2016).

34. 34.

Knill, E. et al. Randomized benchmarking of quantum gates. *Phys. Rev. A* **77**, 012307 (2008).

35. 35.

Magesan, E. et al. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Phys. Rev. Lett.* **109**, 080505 (2012).

36. 36.

Xue, X. et al. CMOS-based cryogenic control of silicon quantum circuits. *Nature* **593**, 205–210 (2021).

37. 37.

Xue, X. et al. Computing with spin qubits at the surface code error threshold. Preprint at <https://arxiv.org/abs/2107.00628> (2021).

38. 38.

Mądzik, M. T. et al. Precision tomography of a three-qubit electron-nuclear quantum processor in silicon. Preprint at <https://arxiv.org/abs/2106.03082> (2021).

39. 39.

Noiri, A. et al. Radio-frequency-detected fast charge sensing in undoped silicon quantum dots. *Nano Lett.* **20**, 947–952 (2020).

40. 40.

Connors, E. J., Nelson, J. & Nichol, J. M. Rapid high-fidelity spin state readout in Si/SiGe quantum dots via radio-frequency reflectometry. *Phys. Rev. Appl.* **13**, 024019 (2020).

41. 41.

Takeda, K. et al. Quantum tomography of an entangled three-qubit state in silicon. *Nat. Nanotechnol.* **16**, 965–969 (2021).

42. 42.

Muhonen, J. T. et al. Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking. *J. Phys. Condens. Matter* **27**, 154205 (2015).

43. 43.

Shulman, M. D. et al. Suppressing qubit dephasing using real-time Hamiltonian estimation. *Nat. Commun.* **5**, 5156 (2014).

44. 44.

Dewes, A. et al. Characterization of a two-transmon processor with individual single-shot qubit readout. *Phys. Rev. Lett.* **108**, 057002 (2012).

Acknowledgements

We thank the Microwave Research Group in Caltech for technical support. This work was supported financially by Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency

(JST) (JPMJCR15N2 and JPMJCR1675), MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) grant numbers JPMXS0118069228, JST Moonshot R&D grant number JPMJMS2065, and JSPS KAKENHI grant numbers 16H02204, 17K14078, 18H01819, 19K14640 and 20H00237. T.N. acknowledges support from JST PRESTO grant number JPMJPR2017.

Author information

Affiliations

1. RIKEN Center for Emergent Matter Science (CEMS), Wako, Japan

Akito Noiri, Kenta Takeda, Takashi Nakajima & Seigo Tarucha

2. RIKEN Center for Quantum Computing (RQC), Wako, Japan

Takashi Kobayashi & Seigo Tarucha

3. QuTech, Delft University of Technology, Delft, The Netherlands

Amir Sammak & Giordano Scappucci

4. Netherlands Organisation for Applied Scientific Research (TNO),
Delft, The Netherlands

Amir Sammak

5. Kavli Institute of Nanoscience, Delft University of Technology, Delft,
The Netherlands

Giordano Scappucci

Contributions

A.N. and K.T. fabricated the device and performed the measurements. T.N. and T.K. contributed the data acquisition and discussed the results. A.S. and G.S. developed and supplied the ^{28}Si /silicon-germanium

heterostructure. A.N. wrote the manuscript with inputs from all co-authors. S.T. supervised the project.

Corresponding authors

Correspondence to [Akito Noiri](#) or [Seigo Tarucha](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Additional information

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

Extended data figures and tables

[Extended Data Fig. 1 Detuning dependence of EDSR spectra.](#)

a, Stability diagram around the (1,1) charge state. **b**, Quantum circuit for producing **c**. The microwave frequency of the $\{\rm \pi\}$ CROT on Q_1 is varied to measure EDSR spectra. **c**, Detuning dependence of EDSR spectra of Q_1 . The detuning axis and its origin are shown as the white arrow and square in **a**. Three black symbols show the conditions where the dephasing times $\{T\}_{2,1}\downarrow^{\ast}$ shown in **d–f** are measured. **d–f**, Ramsey fringes of Q_1 when Q_2 is spin-down measured at

the detuning $\Delta = -0.009$ V (**d**), 0 V (**e**), and 0.009 V (**f**). The integration time is 87 s for all of the traces. The errors in $(\langle T \rangle_{2,1} \downarrow)^{\text{ast}}$ represent the estimated standard errors for the best-fit values. We observe longer (shorter) $(\langle T \rangle_{2,1} \downarrow)^{\text{ast}}$ when the slope of the EDSR frequency against the detuning is smaller (larger), indicating the detuning charge noise limits $(\langle T \rangle_{2,1} \downarrow)^{\text{ast}}$ at the charge-symmetry point where a finite slope exists due to the micromagnet-induced gradient field. A similar tendency is also observed in all the $(\langle T \rangle_{2,1} \downarrow)^{\text{ast}}$.

Extended Data Fig. 2 Qubits characterizations.

a, b, Sequences to measure spin relaxation times for Q_1 when Q_2 is spin-down, $(\langle T \rangle_{1,1} \downarrow)$ (**a**) and -up, $(\langle T \rangle_{1,1} \uparrow)$ (**b**). **c**, Spin-up probability as a function of the wait time. All of the traces do not show a decaying property indicating that spin relaxation is negligible for both qubits. The purple (magenta) curve is obtained using the sequence shown in **a** (**b**). The roles of Q_1 and Q_2 are swapped to measure the data for Q_2 . Each trace is offset by 0.45 for clarity. All of the measurements are performed with $J=18.85$ MHz and $f_{\text{rm}}(R)=4.867$ MHz. **d, e**, Ramsey sequences to measure dephasing times for Q_1 , $(\langle T \rangle_{2,1} \downarrow)^{\text{ast}}$ and $(\langle T \rangle_{2,1} \uparrow)^{\text{ast}}$. **f**, Ramsey fringes of Q_1 and Q_2 fitted with Gaussian decaying oscillation functions. The integration time is 87 s for all of the traces. The errors represent the estimated standard errors for the best-fit values. Each trace is offset by 0.6 for clarity. **g, h**, Echo sequences to measure echo times for Q_1 , $(\langle T \rangle_{2,1} \downarrow)^{\text{echo}}$ and $(\langle T \rangle_{2,1} \uparrow)^{\text{echo}}$. The phase of the final $(\pi/2)$ rotation is varied and the amplitude of the measured oscillation as a function of the phase is plotted in **i**. **i**, Echo amplitudes as a function of the evolution time. The exponent of the decay is $(1.5, 1.2, 1.8)$, and (1.6) for $(\langle T \rangle_{2,1} \downarrow)^{\text{echo}}$, $(\langle T \rangle_{2,1} \uparrow)^{\text{echo}}$, $(\langle T \rangle_{2,2} \downarrow)^{\text{echo}}$, and $(\langle T \rangle_{2,2} \uparrow)^{\text{echo}}$. The errors represent the estimated standard errors for the best-fit values. Each trace is offset by 0.2 for clarity. **j, k**, Measurement of Rabi decay time for Q_1 ,

$(\{T\}_{2,1}\downarrow \{\rm{Rabi}\})$, and $(\{T\}_{2,1}\uparrow \{\rm{Rabi}\})$). We measure Rabi oscillations by varying microwave burst time $\{t\}_{\{\rm{burst}\}}$ from $0.01\mu s$ to $0.41\mu s$ with a separation of $0.01\mu s$. Rabi oscillations for longer $\{t\}_{\{\rm{burst}\}}$ (offset by 20 , 40 , and $80\mu s$) are also measured and the amplitudes of the oscillations are plotted in **I. I**, Rabi oscillation amplitude as a function of the microwave burst time with decaying fits. The decay follows $(R_m\sigma)(t)=\exp(-t/T_{2,m}\sigma)^{\{\rm{Rabi}\}}W(t))$ where $(W(t)=(1+t^2/\{f\}_{\{\rm{R}\}})\{(T_{2,m}\sigma)^{\{\rm{ast}\}}\}^2)^{-1/4})$ represents the effect of dephasing³². From the fit, we extract the Rabi decay during a $(\{\pi\}/2)$ CROT as $(D_m\sigma)=\{R_m\sigma\}(t=1/(4\{f\}_{\{\rm{R}\}}))$. The errors represent the estimated standard errors for the best-fit values. Each trace is offset by 0.5 for clarity.

Extended Data Fig. 3 Single-tone single-qubit gate performance.

a, b, Quantum circuits of single-tone single-qubit Clifford-based randomized benchmarking for Q_1 when Q_2 is spin-down (**a**) and -up (**b**). **c**, Single-tone single-qubit primitive gate fidelities $(F_{\{\rm{p}\},m}\sigma)$ assessed by the Clifford-based randomized benchmarking. The purple (magenta) curve is obtained using the sequence shown in **a** (**b**). The roles of Q_1 and Q_2 are swapped to measure the data for Q_2 . $(f\{\rm{R}\})=4.867$ MHz and $(J=18.85)$ MHz ($=\sqrt{15}$) $\{f\}_{\{\rm{R}\}}$ are used. Each trace is offset by 0.15 for clarity. The uncertainty in the gate fidelities are obtained by a Monte Carlo method⁴. The obtained fidelities are consistent with those obtained in Fig. **2c** as $(F_{\{\rm{p}\},m}\approx F_{\{\rm{p}\},m}\downarrow \{F\}_{\{\rm{p}\},m}\uparrow)$. **d**, Rabi frequency dependence of single-tone single-qubit primitive gate infidelities. Since the control qubit state is fixed in this measurement, the off-resonant rotation does not matter so that $(f\{\rm{R}\})$ can be varied under a fixed (J) of (32.0) MHz. Therefore, the impact of $(f\{\rm{R}\})$ on the single-qubit gate performance is assessed without involving the effect of (J) . We find that the fidelities depend on $(f\{\rm{R}\})$ and the best values are

obtained at $\{f\}_{\rm R}=2$ – 5 MHz. Around the best condition, the fidelities are uniformly high suggesting that the fidelity is mostly limited by pulse imperfections and calibration errors rather than dephasing and Rabi decay effects. The uncertainty in the gate fidelities are obtained by a Monte Carlo method⁴.

Extended Data Fig. 4 Two-qubit gate fidelity extraction.

a, Number of Clifford gates n dependence of the projection state probability $\langle \{P\} \rangle_{\uparrow\uparrow}$ ^{4,16}. The ideal final state is spin-up for both qubits. To extract gate fidelity, we need to measure the saturation value of $\langle \{P\} \rangle_{\uparrow\uparrow}$ with a large n (Methods). The uncertainty in the gate fidelity is obtained by a Monte Carlo method⁴. **b**, Gate fidelity extraction from the sequence fidelity $\langle F \rangle_{\rm rm\{t\}}$. In addition to the data in **a**, we measure another data set where the final ideal state is spin-down for both qubits and then obtain $\langle F \rangle_{\rm rm\{t\}}$ as shown in blue (Methods). The saturation value of $\langle F \rangle_{\rm rm\{t\}}$ is almost zero ($\langle F \rangle_{\rm rm\{t\}}(271)=-0.007$) as expected. Gate fidelity extraction using only the data up to $n=62$ is shown in red. The uncertainty in the gate fidelities are obtained by a Monte Carlo method⁴. The trace is offset by 0.1 for clarity. The obtained gate fidelities agree well with that obtained in the standard protocol in **a**. The uncertainty in the fidelity is larger in **a** due to the uncertainty of the saturation value of $\langle \{P\} \rangle_{\uparrow\uparrow}$. $\{f\}_{\rm R}=5.732$ MHz and $J=22.2$ MHz are used.

Extended Data Fig. 5 Estimation of two-qubit primitive gate infidelity by resonance frequency noise.

a, Time dependence of $\Delta J/2=(\Delta f_1\downarrow-\Delta f_1\downarrow)/2$ (blue), $(\Delta f_1\downarrow+\Delta f_1\downarrow)/2$ (purple), and $(\Delta f_2\downarrow+\Delta f_2\downarrow)/2$ (orange) extracted from repeated Ramsey fringe measurements (Methods). J is fixed at 18.85 MHz. Each trace is offset by 0.25 MHz for clarity. Single-qubit frequency noises (Δf_1 and Δf_2) are larger than

that of the exchange noise $\langle \Delta J/2 \rangle$. **b**, Simulation of a two-qubit primitive gate infidelity by the frequency noises obtained in **a** (Methods). **c**, Similar to **b** but the case with inserting an idle time for both qubits to remove the controlled-phase accumulation during the CROT when switching $\langle J \rangle$ on and off^{18,31}.

Extended Data Fig. 6 Detuning dependence of the two-qubit gate performance.

a, Detuning dependence of $\langle J \rangle$. $\langle J \rangle$ at the charge-symmetry point (detuning $\langle =0 \rangle$ mV) is $\langle 18.85 \rangle$ MHz. **b**, Detuning dependence of the two-qubit primitive gate fidelity $\langle \{F\}_{\{\{rm{p}\}\}} \rangle$ (indigo circles) and the Rabi decay during the $\langle \{\{rm{\pi}\}/2 \} \rangle$ CROT (colored squares) obtained similarly to Fig. 1f. Around the charge-symmetry point, we reproducibly obtain $\langle \{F\}_{\{\{rm{p}\}\}} \rangle$ higher than $\langle 99 \rangle\%$. In large positive and negative detuning, $\langle \{F\}_{\{\{rm{p}\}\}} \rangle$ sharply drops mainly due to the fast Rabi decay. The uncertainty in the gate fidelity is obtained by a Monte Carlo method⁴. The errors in the Rabi decay represent the estimated standard errors for the best-fit values.

Extended Data Fig. 7 Measurement error calibration in state tomography.

Typical joint probabilities measured with preparing $\langle \uparrow \uparrow \rangle$, $\langle \tilde{\{\downarrow \uparrow \downarrow \}} \rangle$, $\langle \tilde{\{\downarrow \downarrow \uparrow \}} \rangle$, and $\langle \{\downarrow \downarrow \downarrow \} \rangle$. At $\langle J \rangle = 18.85$ MHz, $\langle \tilde{\{\downarrow \downarrow \uparrow \}} \rangle = 0.9995 \langle \downarrow \downarrow \uparrow \rangle + 0.0310 \langle \uparrow \uparrow \downarrow \rangle$.

Extended Data Fig. 8 Output state of Deutsch–Jozsa algorithm and Grover search algorithm.

a–c, Real part of the measured density matrix for the final output states for $\langle \{f\}_0 \rangle$ (**a**), $\langle \{f\}_1 \rangle$ (**b**), and $\langle \{f\}_3 \rangle$ (**c**) in the Deutsch–Jozsa algorithm (Fig. 4a). **d–f**, Real part of the measured density matrix for the

final output states for $\left|f_{10}\right\rangle$ (**d**), $\left|f_{01}\right\rangle$ (**e**), and $\left|f_{00}\right\rangle$ (**f**) in the Grover search algorithm (Fig. 4b). The absolute values of the matrix elements for the imaginary parts are less than $|0.055|$ (**a**), $|0.056|$ (**b**), $|0.040|$ (**c**), $|0.111|$ (**d**), $|0.072|$ (**e**), and $|0.081|$ (**f**). The uncertainty in the state fidelities $|F|$ are obtained by a Monte Carlo method^{16,19,41}.

Extended Data Fig. 9 Bell state tomography.

a, Quantum circuit for the Bell state tomography. After the first $\{\text{rm}\{\pi}\}/2$ rotation, $\{\text{rm}\{Z\}\}\text{text}-\{\text{rm}\{CNOT\}\}_2$ and $\{-\{\text{rm}\{Z\}\}_2/2$ $\{\text{rm}\{Z\}\}_2/2$) are applied for **b** (**c**), $\{\text{rm}\{CNOT\}\}_2$ and $\{\text{rm}\{Z\}\}_2/2$) $\{-\{\text{rm}\{Z\}\}_2/2$) are applied for **d** (**e**). $\{I\}, \{X\}/2, \{Y\}/2, \{Z\}$ and $\{X\}$ acting on both qubits at the end change the measurement axis to implement the state tomography (Methods). **b–e**, Real part of the measured density matrix for the prepared Bell states for $\{\Phi^-\}$ (**b**), $\{\Phi^+\}$ (**c**), $\{\Psi^-\}$ (**d**), and $\{\Psi^+\}$ (**e**), respectively. The absolute values of the matrix elements for the imaginary parts are less than $|0.038|$ (**b**), $|0.093|$ (**c**), $|0.100|$ (**d**), $|0.113|$ (**e**). The uncertainty in the state fidelities $|F|$ are obtained by a Monte Carlo method^{16,19,41}.

Supplementary information

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Noiri, A., Takeda, K., Nakajima, T. *et al.* Fast universal quantum gate above the fault-tolerance threshold in silicon. *Nature* **601**, 338–342 (2022).

<https://doi.org/10.1038/s41586-021-04182-y>

- Received: 04 August 2021
- Accepted: 26 October 2021
- Published: 19 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04182-y>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [Precision tomography of a three-qubit donor quantum processor in silicon](#)
 - Mateusz T. MÄ?dzik
 - Serwan Asaad
 - Andrea Morello

Nature (2022)

[Silicon qubits move a step closer to achieving error correction](#)

- Ada Warren
- Sophia E. Economou

News & Views 19 Jan 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04182-y>.

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

- Article
- Open Access
- [Published: 19 January 2022](#)

Quantum logic with spin qubits crossing the surface code threshold

- [Xiao Xue^{1,2}](#),
- [Maximilian Russ](#) [ORCID: orcid.org/0000-0001-9775-0323^{1,2}](#),
- [Nodar Samkharadze^{1,3}](#),
- [Brennan Undseth^{1,2}](#),
- [Amir Sammak^{1,3}](#),
- [Giordano Scappucci](#) [ORCID: orcid.org/0000-0003-2512-0079^{1,2}](#) &
- [Lieven M. K. Vandersypen](#) [ORCID: orcid.org/0000-0003-4346-7877^{1,2}](#)

Nature volume **601**, pages 343–347 (2022)

- 5664 Accesses
- 223 Altmetric
- [Metrics details](#)

Subjects

- [Quantum dots](#)
- [Quantum information](#)

Abstract

High-fidelity control of quantum bits is paramount for the reliable execution of quantum algorithms and for achieving fault tolerance—the ability to correct errors faster than they occur¹. The central requirement for fault tolerance is expressed in terms of an error threshold. Whereas the actual threshold depends on many details, a common target is the approximately 1% error threshold of the well-known surface code^{2,3}. Reaching two-qubit gate fidelities above 99% has been a long-standing major goal for semiconductor spin qubits. These qubits are promising for scaling, as they can

leverage advanced semiconductor technology⁴. Here we report a spin-based quantum processor in silicon with single-qubit and two-qubit gate fidelities, all of which are above 99.5%, extracted from gate-set tomography. The average single-qubit gate fidelities remain above 99% when including crosstalk and idling errors on the neighbouring qubit. Using this high-fidelity gate set, we execute the demanding task of calculating molecular ground-state energies using a variational quantum eigensolver algorithm⁵. Having surpassed the 99% barrier for the two-qubit gate fidelity, semiconductor qubits are well positioned on the path to fault tolerance and to possible applications in the era of noisy intermediate-scale quantum devices.

[Download PDF](#)

Main

Quantum computation involves the execution of a large number of elementary operations that take a qubit register through the steps of a quantum algorithm⁶. A major challenge is to implement these operations with sufficient accuracy to arrive at a reliable outcome, even in the presence of decoherence and other error sources. The higher the accuracy, or fidelity, of the operations, the higher the likelihood that near-term applications for quantum computers come within reach⁷. Furthermore, for most presently known algorithms, the number of operations that must be concatenated will unavoidably lead to excessive accumulation of errors, and these errors must be removed using quantum error correction¹. Correcting quantum errors faster than they occur is possible when the error probability per operation is below a certain threshold, known as the fault-tolerance threshold. For the widely considered surface code, for instance, the fault-tolerance threshold is between 0.6% and 1%, under certain assumptions, albeit at the cost of a large redundancy in the number of physical qubits^{2,3}.

Among all the candidate platforms, electron spins in semiconductor quantum dots have advantages, such as their long coherence times⁸, small footprint⁹, the potential for scaling up¹⁰ and the compatibility with advanced semiconductor manufacturing technology⁴. Single-qubit operations of spin qubits in quantum dots achieve fidelities of 99.9% (refs. [11,12](#)) but the two-qubit gate fidelities reported vary from 92% to 98% (refs. [13,14](#)). This has limited the two-qubit Bell-state fidelities to 94% (ref. [15](#)) and quantum algorithms implemented with spin qubits gave only coarsely accurate outcomes [16,17](#). Pushing the two-qubit gate fidelity well beyond 99% requires not only low charge-noise levels and the elimination of nuclear spins by isotopic enrichment but also careful Hamiltonian engineering.

In this paper, using a precisely engineered two-qubit interaction Hamiltonian, we report the demonstration of single-qubit and two-qubit gates with fidelities above 99.5%. We use gate-set tomography (GST) not only to characterize the gates and to quantify the fidelity but also to improve the gate calibration. The high-fidelity gates allow us to compute the dissociation energy of molecular hydrogen with a variational quantum eigensolver (VQE) algorithm, reaching an accuracy for the dissociation energy of around 20 mHa, limited by readout errors.

We use a gate-defined double quantum dot in an isotopically enriched $^{28}\text{Si}/\text{SiGe}$ heterostructure¹⁷ (Fig. 1a), with each dot occupied by a single electron (see [Methods](#)). The spin states of the electrons serve as qubits. The spin states are measured with the help of a sensing quantum dot (SQD), which is capacitively coupled to the qubit dots¹⁸. A micromagnet on top of the device provides a magnetic field gradient enabling electric-dipole spin resonance¹⁹ and separates the resonance frequencies of the qubits in the presence of an external magnetic field (~ 320 mT) to 11.993 GHz (Q_1) and 11.890 GHz (Q_2). Single-qubit X and Y gates are implemented by frequency-multiplexed microwave signals applied to gate MW and virtual Z gates are implemented by a phase update of the reference frame²⁰. The plunger gates (LP and RP) control the chemical potentials of the quantum dots.

Fig. 1: Two-qubit device and symmetry operating point.

 [figure 1](#)



a, Scanning electron microscopy images of a device similar to that used here, showing the quantum dot gate pattern and the micromagnet on top (the device used in the experiment has an additional screening gate above the fine gates¹⁷). The scale bar in

the left panel denotes 500 nm. The scale bar in the right panel denotes 100 nm. **b**, Control paths for determining the symmetry operation point in the charge-stability diagram. (M, N) represent the number of electrons in the dots underneath the tip of LP and RP, respectively. a.u., arbitrary units. **c**, Pulse sequence schematic of a decoupled controlled-phase operation interleaved in a Ramsey interference sequence on Q_1 . **d**, Spin-up probability of Q_1 after the Ramsey sequence in **c**, as a function of the detuning in the double-dot potential and the total duration of the barrier voltage pulses.

The native two-qubit gate for spin qubits uses the exchange interaction^{21,22}, originating from the wave-function overlap of electrons in neighbouring dots. This selectively shifts the energy of the antiparallel spin states and, thus, enables an electrically pulsed adiabatic conditional Z (CZ) gate^{8,16,23}. The barrier gate (B) controls the tunnel coupling between the dots, allowing the precise tuning of the exchange coupling from <100 kHz to 20 MHz. To minimize the sensitivity to charge noise, we activate the exchange coupling while avoiding a tilt in the double-dot potential^{24,25} (Fig. 1a). This symmetric condition can be determined accurately by decoupled adiabatic exchange pulses inside a Ramsey sequence (Fig. 1c, d). The tunnel barrier is controlled by simultaneously pulsing gate B and compensating LP and RP to avoid shifts in the electrochemical potentials²⁴, constituting a virtual barrier gate. The detuning between quantum dots is controlled by additional offsets to the LP and RP pulses in opposite directions. As the decoupling pulses remove additional single-qubit phase accumulation from electron movement in the magnetic field gradient, the spin-up probability of Q_1 results in a symmetric chevron pattern, with the symmetry point at the centre (Fig. 1d).

Among the various quantum benchmarking techniques, quantum process tomography (QPT) is designed to reconstruct all details in a target process⁶. Owing to the susceptibility of QPT to state preparation and measurement (SPAM) errors, self-consistent benchmarking techniques such as GST²⁶ and alternative techniques such as randomized benchmarking²⁷ have been developed. In contrast to randomized benchmarking, GST inherits the advantage of QPT in that it reports the detailed process, which allows us to isolate Hamiltonian errors from stochastic errors and to correct for such errors in the control signals (Extended Data Fig. 5). In addition, GST accounts for gate-dependent errors. We benchmark the fidelities of a universal gate set using GST^{26,28} (Fig. 2a). The gate set we choose contains an idle gate (I), sequentially operated single-qubit $\pi/2$ rotations about the (\hat{x}) and (\hat{y}) axes for each qubit ($\{\{\rm{rm}X\}\}_{\{\{\rm{rm}Q\}\}_{\{1\}}}, \{\{\rm{rm}Y\}\}_{\{\{\rm{rm}Q\}\}_{\{1\}}}, \{\{\rm{rm}X\}\}_{\{\{\rm{rm}Q\}\}_{\{2\}}}, \{\{\rm{rm}Y\}\}_{\{\{\rm{rm}Q\}\}_{\{2\}}}$) and a two-qubit controlled-phase (CZ) gate. A total of 36 fiducial sequences containing $\{\{\rm{rm}n\}\{\rm{rm}u\}\{\rm{rm}l\}\{\rm{rm}i\}, \{\{\rm{rm}X\}\}_{\{\{\rm{rm}Q\}\}_{\{i\}}}\}^{\{n=1,2,3\}}, \{\{\rm{rm}Y\}\}_{\{\{\rm{rm}Q\}\}_{\{j\}}}\}^{\{n=1,3\}}$ on each qubit, where null (unlike the

idle gate) has no waiting time, are used to tomographically measure the two-qubit state. These fiducials are interleaved by germ sequences and their powers up to a sequence depth of 16. Germs are short sequences of gates taken from the universal gate set (see [Methods](#)). They are repetitively executed to amplify different types of gate errors in the gate set, such that SPAM errors can be isolated. GST allows using a maximum-likelihood estimator to compute completely positive and trace-preserving process matrices for each element of the gate set^{[6](#)}. The gate fidelity can be calculated by comparing the measured process using the Pauli transfer matrix (PTM), $\langle \{ \mathcal{M} \}_\{\exp\} \rangle$, with the ideal PTM, $\langle \{ \{ \mathcal{M} \}_\{\rm{ideal}\} \} \rangle$, $\langle F \{ \{ \mathcal{M} \}_\{\rm{gate}\} \} \rangle = \langle \{ \rm{Tr} \} (\{ \{ \mathcal{M} \}_\{\exp\}^{-1} \{ \mathcal{M} \}_\{\rm{ideal}\} \}) + d \rangle [d(d+1)]$, where d is the dimension of the Hilbert space. These process matrices provide a detailed error diagnosis of the gate set, allowing for efficient feedback calibration^{[29](#)} (Fig. [2a](#)). Analysing the error generator $\langle \{ \mathcal{L} \} = \log (\{ \{ \mathcal{M} \}_\{\exp\} \{ \{ \mathcal{M} \}_\{\rm{ideal}\} \}^{-1}) \rangle$ provides easy access to information. For example, coherent Hamiltonian errors can be isolated from incoherent stochastic errors and single-qubit errors can be isolated from each other and from two-qubit errors^{[30](#)}.

Fig. 2: Gate-set tomography and single-qubit gate.

 **figure 2**

a, Workflow of the GST experiment. Coloured blocks show the input and output fiducial sequences (Fid_i and Fid_o , orange) and the germ sequences (green). A few examples of single-qubit germ sequences are listed. The outcome is used to adjust pulse parameters in the next run. **b, c**, PTMs of $\langle \{\{\text{rm}\{X\}}\} \rangle \langle \{\{\text{rm}\{Q\}}\} \rangle \langle \{1\} \rangle$

and $\langle \{\{\rm{rm}\{Y\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{1\} \rangle$) in the subspace of Q_1 . The red (blue) bars are theoretically +1 (-1) and are measured to be positive (negative). The brown (green) bars are theoretically 0 (0) but measured to be positive (negative). P_{in} and P_{out} are the input and output operators, respectively. **d**, Experimentally measured PTM of $\langle \{\{\rm{rm}\{Y\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{1\} \rangle \otimes \langle \{\{\rm{rm}\{I\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{2\} \rangle$) in the complete two-qubit space. The colour code is the same as in **b, c**.

Figure 2b,c shows the reduced PTMs of $\langle \{\{\rm{rm}\{X\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{1\} \rangle$ and $\langle \{\{\rm{rm}\{Y\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{1\} \rangle$ operations in the Q_1 subspace and Fig. 2d shows the full PTM of $\langle \{\{\rm{rm}\{Y\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{1\} \rangle$ in two-qubit space ($\langle \{\{\rm{rm}\{Y\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{1\} \rangle \otimes \langle \{\{\rm{rm}\{I\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{2\} \rangle$) containing additional errors from decoherence and crosstalk on Q_2 while operating Q_1 (see Extended Data Figs. 1 and 2 for other PTMs) and from unintentional entanglement due to a residual exchange interaction. The average single-qubit gate fidelity is 99.72% in the single-qubit subspace ($\langle \{\{\rm{rm}\{X\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{1\} \rangle$): 99.68%; $\langle \{\{\rm{rm}\{Y\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{1\} \rangle$): 99.73%; $\langle \{\{\rm{rm}\{X\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{2\} \rangle$): 99.61%; $\langle \{\{\rm{rm}\{Y\}}\} \rangle \langle \{\{\rm{rm}\{Q\}}\} \rangle \langle \{2\} \rangle$): 99.87%; see Extended Data Fig. 2 for all error bars). A metric that is rarely reported is the single-qubit gate fidelity in the full two-qubit space, here 99.16% on average (see [Methods](#) and Extended Data Fig. 1). These results highlight that single-qubit benchmarking is not sufficient to identify all errors occurring during single-qubit operations. By analysing the error generators, we find that errors from uncorrelated dephasing of the idling qubit dominate the drop in single-qubit gate fidelity when characterized in the two-qubit space. Coherent, microwave-induced phase shifts—the main source of crosstalk errors—have been corrected by applying a compensating phase gate to the idling qubit (Extended Data Fig. 4). The elimination of idling errors and other crosstalk errors from the microwave drive, such as through heating effects, will be a crucial step to improve the quality of the single-qubit operations further.

For a high-fidelity adiabatic CZ gate, precise control of the exchange coupling, J , between the two qubits is required. Specifically, in order to avoid unintended state transitions due to non-adiabatic dynamics, we must be able to carefully shape the envelope of J . We characterize J over a wide range using a Ramsey sequence interleaved by a virtual barrier pulse with incremental amplitude v_B . Figure 3a shows the measured frequency shift of each qubit as functions of the barrier pulse amplitude and the state of the other qubit. The exchange interaction is modelled to be exponentially dependent on the barrier pulse amplitude $\langle J(v) \rangle \langle \{\rm{rm}\{B\}\} \rangle \propto e^{\alpha v}$ (refs. [31,32](#)). The micromagnet-induced single-qubit frequency shifts are approximated by linear functions within the voltage window of the CZ gate in the numerical simulations. By fitting the measured datasets simultaneously to theoretical models (see [Methods](#)), J can be extracted very precisely

as the difference between the two conditional frequencies of each qubit^{16,33} (Fig. 3b). The barrier pulse $\{v\}_{\{\rm{B}\}} \propto \log(\{A\}_{\{v\}_{\{\rm{B}\}}})(1 - \cos(2\pi\{g\}_{\{t\}_{\{\rm{B}\}}}\{a\}_{\{t\}_{\{\rm{B}\}}}))$ (Fig. 3d) compensates the exponential dependence such that $J \propto (1 - \cos(2\pi t/t_{\text{gate}}))$ follows a cosine window function, which ensures good adiabaticity³⁴ (Fig. 3e). In addition, the virtual gates are calibrated such that the symmetric operation point is maintained for each barrier setting, minimizing the influence of charge noise via the double-dot detuning. The most relevant remaining noise sources include charge noise, affecting J through fluctuations in the virtual barrier gate δv_B , and fluctuating qubit frequencies $\langle \delta f \rangle_{\{Q\}_1}$ and $\langle \delta f \rangle_{\{Q\}_2}$ from charge noise entering through artificial spin-orbit coupling from the micromagnet and residual nuclear spin noise coupling through the hyperfine interaction. By analysing the decay of the Ramsey oscillations at each transition frequency, individual dephasing times $\langle T_2^{\text{ast}} \rangle$ can be extracted and, from there, also δv_B , $\langle \delta f \rangle_{\{Q\}_1}$ and $\langle \delta f \rangle_{\{Q\}_2}$ (Fig. 3c).

Fig. 3: Hamiltonian engineering of exchange interaction.

 **figure 3**

a, Frequency detuning of each qubit conditional on the state of the other qubit as a function of barrier pulse amplitude. The horizontal axis shows the real voltage applied to gate B. **b**, Exchange strength as a function of barrier pulse amplitude. The data are extracted directly from **a**. **c**, $\langle \{T\}_2^{\text{ast}} \rangle$ of each qubit conditional on the state of the other qubit as a function of barrier pulse amplitude (same colour code as in **a**). Each data point is averaged for about 8 min. By fitting the $\langle \{T\}_2^{\text{ast}} \rangle$ values to a quasistatic noise model (solid lines, see [Methods](#)), the low-frequency amplitudes of the fluctuations are estimated as $\langle \delta f \rangle_{\{Q\}_1} = 11 \text{ mHz}$, $\langle \delta f \rangle_{\{H\}_1} = 24 \text{ mHz}$ and $\delta v_B = 0.4 \text{ mV}$. **d**, Shape of the barrier pulse, designed to achieve a high-fidelity CZ gate. **e**, The cosine-shaped J envelope seen by the qubits during the pulse shown in **d**.

Figure 4a shows an example GST pulse sequence that contains twice in a row the germ $\{[\{\text{rm}\{C\}\} \{\text{rm}\{Z\}\}, \{\{\text{rm}\{X\}\}\}_{\{\{\text{rm}\{Q\}\}\}_{\{2\}}}], \{\{\text{rm}\{Y\}\}\}_{\{\{\text{rm}\{Q\}\}\}_{\{1\}}}, \{\text{rm}\{C\}\} \{\text{rm}\{Z\}\}, \{\{\text{rm}\{Y\}\}\}_{\{\{\{\text{rm}\{Q\}\}\}_{\{2\}}}, \{\{\text{rm}\{X\}\}\}_{\{\{\{\text{rm}\{Q\}\}\}_{\{1\}}}\}$. The PTM of the CZ gate obtained from GST is shown in Fig. 4b. Using the detailed information from the error generator to fine-tune the calibration parameters, we can achieve a CZ fidelity of $99.65 \pm 0.15\%$ (Extended Data Figs. 4 and 5). Error bars included here and elsewhere are the $2\sigma \approx 95\%$ confidence intervals computed using the Hessian of the loglikelihood function³⁵. The CZ error generator reveals that, at this point, incoherent errors dominate. The virtual barrier gate technique used here efficiently suppresses crosstalk errors during two-qubit gates. Therefore, we expect the CZ fidelity to be mostly affected by dephasing errors of idling qubits in a larger space, which can be corrected for using decoupling pulses. From the obtained PTMs, we can numerically estimate Bell-state fidelities by multiplications of the PTMs necessary to construct the corresponding state, giving an estimate of 97.75%–98.42%, neglecting SPAM errors, for the four Bell states (Fig. 4c and Extended Data Fig. 3).

Fig. 4: High-fidelity two-qubit gate.



a, A sequence of pulses generated by the arbitrary waveform generators in an example GST sequence. The purple waveforms show the in-phase component of X/Y gates. The CZ gate is indicated by the orange pulse of gate B and the blue and red compensation pulses of gate LP and gate RP. **b**, Experimentally determined PTM of a CZ gate. The colour code is the same as in Fig. 2. **c**, Left, the quantum circuit used to reconstruct the Bell state $\langle|\{\Psi\}\rangle^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ based on the corresponding PTMs. Right, the real part of the reconstructed density matrix of the $|\Psi^+\rangle$ state. The colour code is the same as in Fig. 2, except that red (blue) bars here are theoretically +0.5 (-0.5).

Next, we use the high-fidelity gate set in the context of an actual application, in order to provide a quantitative benchmark for future work under realistic conditions.

Specifically, we implement a VQE algorithm to compute the ground-state energy of molecular hydrogen (H_2) (Fig. 5a). In a VQE algorithm, a quantum processor is used to implement a classically inefficient subroutine (see [Methods](#) and Extended Data Fig. 6). The second quantized H_2 Hamiltonian can be mapped onto two qubits under the Bravyi–Kitaev (BK) transformation $H = h_0II + h_1ZI + h_2IZ + h_3ZZ + h_4XX + h_5YY$.

Here I, X, Y and Z are Pauli operators, for example, ZI is shorthand for $Z \otimes I$, and the coefficients h_0-h_5 are classically computable functions of the internuclear distance, R . Figure 5b shows the schematic of the VQE algorithm and its circuit implementation for a H_2 molecule. The qubit is initialized in $|01\rangle$, which represents double occupation of the lowest molecular orbital, corresponding to the Hartree–Fock (HF) ground state. A parameterized ansatz state is then prepared by considering single and double excitation, which, after the BK transformation, yields $\langle|\psi(\theta)\rangle| = \{\{\rm{e}\}\}^{\{-\{\rm{i}\}\}}\theta\{\rm{XY}\}|01\rangle$, with θ the parameter to variationally optimize. By performing partial tomography on the ansatz state with an initial guess θ_0 , the expectation value of the Hamiltonian for $|\psi(\theta_0)\rangle$ can be calculated. A classical computer can efficiently compute the next guess θ_1 as the new input for the quantum computer. This loop is iterated until the result converges. For a H_2 molecule, there is only one parameter θ to optimize, thus, a scan of the entire parameter range of 2π with finite samples is sufficient to interpolate the smoothly changing measured expectation values. This emulates a real variational algorithm, where θ can be estimated to arbitrary precision by increasing the number of repetitions to suppress statistical fluctuations³⁶. Figure 5c shows the partial tomography result after normalization of the visibility window. The data demonstrate high-quality phase control in the quantum circuits. The deviations in the odd-parity expectation values indicate correlations in the readout of the two qubits³⁷. Figure 5d shows the energy curves of the H_2 molecule from both theory³⁸ and the VQE experiment. We observe a minimum energy at around 0.72 Å and an error of approximately 20 mHa at the theoretical bond length 0.7414 Å, mainly attributed to slow drift in the readout parameters. This accuracy matches the results obtained using superconducting and trapped ion qubits with comparable gate fidelities^{36,39}.

Fig. 5: Variational quantum eigensolver.



a, Lowest two molecular orbitals of a H₂ molecule, formed by the 1s orbitals of two hydrogen atoms. **b**, The quantum circuit to implement the VQE algorithm for a H₂ molecule. The orange block prepares the HF initial state by flipping Q₂. The circuit in green blocks creates the parameterized ansatz state. \(-\{\{\rm{X}\}\}\}_{-}\{\{\{\rm{Q}\}\}\}_{-}\{i\}\) and \(-\{\{\rm{Y}\}\}\}_{-}\{\{\{\rm{Q}\}\}\}_{-}\{j\}\) include virtual Z gates. CNOT gates are compiled as \([-[\{\rm{Y}\}\}_{-}\{\{\{\rm{Q}\}\}\}_{-}2], \{\rm{C}\}\{\rm{Z}\}, \{\{\rm{Y}\}\}_{-}\{\{\{\rm{Q}\}\}\}_{-}2]\). To make use of the high-fidelity CZ gate, such compilation is preferred instead of using a single controlled-phase gate with incremental length for creating the parameterized ansatz state. **c**, Expectation values of the operators in the two-qubit Hamiltonian under BK transformation as a function of θ . Black solid lines show the predicted values. The coloured solid lines are sinusoidal fits to the data (and a constant fit for the case of ZZ). **d**, Potential energy of the H₂ molecule at varying R . The VQE data are normalized to the theoretical energy at large R to directly compare the dissociation energy with the theoretical value. The inset shows the error in the normalized experimental data.

The two-qubit gate with fidelity above 99.5% and single-qubit gate fidelities in the two-qubit gate space above 99% on average place semiconductor spin qubit logic at the error threshold of the surface code. Recently, a two-qubit operation between nuclear spin qubits in silicon, mediated by an electron spin qubit, has been demonstrated to surpass 99% fidelity as well, further highlighting that semiconductor spin qubits offer precise two-qubit logic⁴⁰. Independent studies have shown spin qubit readout with a fidelity above 98% in only a few μs (ref. ⁴¹), with further improvements

underway⁴². Combining high-fidelity initialization, readout and control into a demonstration of fault tolerance poses several key challenges to be overcome. First, sufficiently large and reliable quantum dot arrays must be constructed, with good connectivity between the qubits. Second, the fidelities achieved in small-scale systems must be maintained across such larger systems, which will require reducing idling and crosstalk errors. The same advances will allow us to implement more sophisticated algorithms in the noisy intermediate-scale quantum era, such as solving energies involving excited states of more complex molecules.

Methods

Measurement setup

The measurement setup and device are similar to those used in ref. ¹⁷. We summarize a few key points and all the differences here. The gates LP, RP and B are connected to arbitrary waveform generators (AWGs, Tektronix 5014C) via coaxial cables. The position in the charge-stability diagram of the quantum dots is controlled by voltage pulses applied to LP and RP. Linear combinations of the voltage pulses applied to B, LP and RP are used to control the exchange coupling between the two qubits at the symmetry point. The compensation coefficients are $v_{LP}/v_B = -0.081$ and $v_{RP}/v_B = 0.104$. A vector signal generator (VSG, Keysight E8267D) is connected to gate MW and sends frequency-multiplexed microwave bursts (not necessarily time-multiplexed) to implement electric-dipole spin resonance (EDSR). The VSG has two I/Q input channels, receiving I/Q modulation pulses from two channels of an AWG. I/Q modulation is used to control the frequency, phase and length of the microwave bursts. The current signal of the sensing quantum dot is converted to a voltage signal and recorded by a digitizer card (Spectrum M4i.44), and then converted into 0 or 1 by comparing it to a threshold value.

Two differences between the present setup and that in ref. ¹⁷ are that (1) the programmable mechanical switch is configured such that gate MW is always connected to the VSG and not to the cryo-CMOS control chip and (2) a second AWG of the same model is connected to gate B, with its clock synchronized to the first AWG.

Gate calibration

In the gate set used in this work, $\{\{\rm I\}, \{\rm X\}\}_{\{\{\rm Q\}\}_1}, \{\{\rm Y\}\}_{\{\{\rm Q\}\}_1}, \{\{\rm X\}\}_{\{\{\rm Q\}\}_2}, \{\{\rm Y\}\}_{\{\{\rm Q\}\}_2}, \{\rm C\} \{\rm Z\}$, the duration of the I gate and the CZ gate are set to 100 ns, and we calibrate and keep the amplitudes of the

single-qubit drives fixed and in the linear-response regime, where the Rabi frequency is linearly dependent on the driving amplitude. The envelopes of the single-qubit gates are shaped following a ‘Tukey’ window, as it allows adiabatic single-qubit gates with relatively small amplitudes, thus, avoiding the distortion caused by a nonlinear response. The general Tukey window of length t_p is given by

```

$$W(t,r)=\begin{array}{cc}\frac{1}{2}\left[1-\cos \left(\frac{2\pi r}{p}t\right)\right] & 0\leq t\leq \frac{r}{p}\\ \frac{1}{2}\left(1-\cos \left(\frac{2\pi (t-r)}{p}\right)\right) & \frac{r}{p} < t < \frac{r}{p}-\frac{1}{2}\\ \frac{1}{2}\left(1-\cos \left(\frac{2\pi (t-r)}{p}\right)\right) & \frac{r}{p}-\frac{1}{2} \leq t \leq \frac{r}{p}\end{array}
(1)

```

where $r = 0.5$ for our pulses. Apart from these fixed parameters, there are 11 free parameters that must be calibrated: single-qubit frequencies $\langle\{f\}\rangle_{\{\{\rm{rm}\{Q\}\}\}}_1$ and $\langle\{f\}\rangle_{\{\{\rm{rm}\{Q\}\}\}}_2$, burst lengths for single-qubit gates t_{XY1} and t_{XY2} , phase shifts caused by single-qubit gates on the addressed qubit itself ϕ_{11} and ϕ_{22} , phase shifts caused by single-qubit gates on the unaddressed ‘victim qubit’ ϕ_{12} and ϕ_{21} (ϕ_{12} is the phase shift on Q_1 induced by a gate on Q_2 and similar for ϕ_{21}), the peak amplitude of the CZ gate $\langle\{A\}\rangle_{\{v\}\{\{\rm{rm}\{B\}\}\}}$ and phase shifts caused by the gate voltage pulses used for the CZ gate on the qubits θ_1 and θ_2 (in addition, we absorb into θ_1 and θ_2 the 90° phase shifts needed to transform $\text{diag}(1, i, i, 1)$ into $\text{diag}(1, 1, 1, -1)$).

For single-qubit gates, $\langle\{f\}_{-}\{\{\rm{rm}\{Q\}\}}_{-}\{1\}\rangle$ and $\langle\{f\}_{-}\{\{\rm{rm}\{Q\}\}}_{-}\{2\}\rangle$ are calibrated by standard Ramsey sequences, which are automatically executed every 2 h, at the beginning and in the middle (after 100 times the average of each sequence) of the GST experiment. The EDSR burst times t_{XY1} and t_{XY2} are initially calibrated by an AllXY calibration protocol⁴³. The phases ϕ_{11} , ϕ_{12} , ϕ_{21} and ϕ_{22} are initially calibrated by measuring the phase shift of the victim qubit (Q_1 for ϕ_{11} and ϕ_{21} ; Q_2 for ϕ_{22} and ϕ_{12}) in a Ramsey sequence interleaved by a pair of $\langle[\{\{\rm{rm}\{X\}\}}_{-}\{\{\rm{rm}\{Q\}\}}_{-}\{i\}\}, -\{\{\rm{rm}\{X\}\}}_{-}\{\{\rm{rm}\{Q\}\}}_{-}\{i\}\}]\rangle$ gates on the addressed qubit (Q_1 for ϕ_{11} and ϕ_{12} ; Q_2 for ϕ_{22} and ϕ_{21}) (Extended Data Fig. 4).

The optimal pulse design presented in Fig. 3 gives a rough guidance of the pulse amplitude $\langle \{A\}_v \{rm{B}\} \rangle$. In a more precise calibration of the CZ gate, an optional π -rotation is applied to the control qubit (for example, Q_1) to prepare it into the $|0\rangle$ or $|1\rangle$ state, followed by a Ramsey sequence on the target qubit (Q_2) interleaved by an exchange pulse. The amplitude is precisely tuned to bring Q_2

completely out of phase (by 180°) between the two measurements (Extended Data Fig. 4d,e). The phase θ_2 is determined such that the phase of Q_2 changes by zero (π) when Q_1 is in the state $|0\rangle$ ($|1\rangle$), corresponding to $CZ = \text{diag}(1, 1, 1, -1)$ in the standard basis. The same measurement is then performed again with Q_2 as the control qubit and Q_1 as the target qubit to determine θ_1 (ref. 16).

In such a ‘conventional’ calibration procedure of the CZ gate, we notice that the two qubits experience different conditional phases (Extended Data Fig. 4). We believe that this effect is caused by off-resonant driving from the optional π -rotation on the control qubit. Similar effects can also affect the calibration of the phase crosstalk from single-qubit gates.

This motivates us to use the results from GST as feedback to adjust the gate parameters. The error generators not only describe the total errors of the gates but also distinguish Hamiltonian errors (coherent errors) from stochastic errors (incoherent errors). We use the information on seven different Hamiltonian errors (IX, IY, XI, YI, ZI, IZ and ZZ) of each gate to correct all 11 gate parameters (Extended Data Fig. 5), except $\langle f \rangle_{\{\rm rm\{Q\}\}}_{\{1\}}$ and $\langle f \rangle_{\{\rm rm\{Q\}\}}_{\{2\}}$, for which calibrations using standard Ramsey sequences are sufficient. For single-qubit gates, t_{XY1} and t_{XY2} are adjusted according to the IX, IY, XI and YI errors. The phases ϕ_{11} , ϕ_{12} , ϕ_{21} and ϕ_{22} are adjusted according to the ZI and IZ errors. For the CZ gate, θ_1 and θ_2 are adjusted according to the ZI and IZ errors, and $\langle A \rangle_{\{v\}}_{\{\rm rm\{B\}\}}$ is adjusted according to the ZZ error. The adjusted gates are then used in a new GST experiment.

Theoretical model

In this section, we describe the theoretical model used for the fitting, the pulse optimization and the numerical simulations. The dynamics of two electron spins in the (1,1) charge configuration can be accurately described by an extended Heisenberg model²¹

$$\begin{aligned} H = & g\{\mu\}_{\{\rm rm\{B\}\}}\{\{\bf B\}\}_{\{1\}} \cdot \{\bf S\}_{\{1\}} + g\{\mu\}_{\{\rm rm\{B\}\}}\{\{\bf B\}\}_{\{2\}} \cdot \{\bf S\}_{\{2\}} + hJ(\{\{\bf S\}\}_{\{1\}} \cdot \{\bf S\}_{\{2\}} - \frac{1}{4}), \\ (2) \end{aligned}$$

with $\langle \{\{\bf S\}\}_{\{j\}} \rangle = \langle \{X\}_{\{j\}}, \{Y\}_{\{j\}}, \{Z\}_{\{j\}} \rangle^{\wedge \{\{\rm rm\{T\}\}\}/2}$, where X_j , Y_j and Z_j are the single-qubit Pauli matrices acting on spin $j = 1, 2$, μ_B the Bohr’s magneton, $g \approx 2$ the g -factor in silicon and h is Planck’s constant. The first and second terms describe the interaction of the electron spin in dot 1 and dot 2 with the magnetic

fields $\{\{\bf{B}\}\}_{j}=\{\{B\}_{x,j},0,\{B\}_{z,j}\}\}^{\{\rm{rm}\{T\}\}}$) originating from the externally applied field and the micromagnet. The transverse components $B_{x,j}$ induce spin-flips, thus, single-qubit gates if modulated resonantly via EDSR. For later convenience, we define the resonance frequencies by $\{h\{f\}_{\{Q\}_1}\}=g\{\mu\}_{\{\rm{rm}\{B\}\}}\{B\}_{z,1}\}$ and $\{h\{f\}_{\{Q\}_2}\}=g\{\mu\}_{\{\rm{rm}\{B\}\}}\{B\}_{z,2}\}$, and the energy difference between the qubits $\Delta E_z=g\mu_B(B_{z,2}-B_{z,1})$. The last term in the Hamiltonian of equation (2) describes the exchange interaction J between the spins in neighbouring dots. The exchange interaction originates from the overlap of the wave functions through virtual tunnelling events and is, in general, a nonlinear function of the applied barrier voltage v_B . We note that v_B determines the compensation pulses applied to LP and RP for virtual barrier control. We model J as an exponential function^{31,32}

$$\$ \$ J(v_{\{\rm{rm}\{B\}\}})=J_{\{\rm{res}\}}\{\{\rm{rm}\{e\}\}\}^{2\alpha v_{\{\rm{rm}\{B\}\}}}, \$ \$ \quad (3)$$

where $J_{\rm{res}} \approx 20-100$ kHz is the residual exchange interaction during idle and single-qubit operations and α is the lever arm. In general, the magnetic fields $\{\{\bf{B}\}\}_{j}$ depend on the exact position of the electron. We include this in our model $\{\{B\}_{z,j}\} \rightarrow \{B\}_{z,j}(v_{\{\rm{rm}\{B\}\}})=\{B\}_{z,j}(0)+\{\beta_j\}v_{\{\rm{rm}\{B\}\}}^{\gamma}$ where β_j accounts for the impact of the barrier voltage on the resonance frequency of qubit j . The transition energies described in the main text are now given by diagonalizing the Hamiltonian from equation (2) and computing the energy difference between the eigenstates corresponding to the computational basis states $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ (ref. ⁴⁴). We have

$$\$ \$ h\{f\}_{\{\{\rm{rm}\{Q\}\}\}_1}(\{\{\rm{rm}\{Q\}\}\}_2=|0\rangle)=\{\mathcal{E}\}(|10\rangle)-\{\mathcal{E}\}(|00\rangle), \$ \$ \quad (4)$$

$$\$ \$ h\{f\}_{\{\{\rm{rm}\{Q\}\}\}_1}(\{\{\rm{rm}\{Q\}\}\}_2=|1\rangle)=\{\mathcal{E}\}(|11\rangle)-\{\mathcal{E}\}(|01\rangle), \$ \$ \quad (5)$$

$$\$ \$ h\{f\}_{\{\{\rm{rm}\{Q\}\}\}_2}(\{\{\rm{rm}\{Q\}\}\}_1=|0\rangle)=\{\mathcal{E}\}(|01\rangle)-\{\mathcal{E}\}(|00\rangle), \$ \$ \quad (6)$$

$$\$ \$ h\{f\}_{\{\{\rm{rm}\{Q\}\}\}_2}(\{\{\rm{rm}\{Q\}\}\}_1=|1\rangle)=\{\mathcal{E}\}(|11\rangle)-\{\mathcal{E}\}(|10\rangle), \$ \$ \quad (7)$$

where $\{\mathcal{E}\}(|\xi\rangle)$ denotes the eigenenergy of eigenstate $|\xi\rangle$ and $|0\rangle=|\downarrow\rangle$ is defined by the magnetic field direction.

In the presence of noise, qubits start to lose information. In silicon, charge noise and nuclear noise are the dominating sources of noise. In the absence of two-qubit coupling and correlated charge noise, both qubits decohere largely independently of each other, giving rise to a decoherence time set by the interaction with the nuclear spins and charge noise coupling to the qubit via intrinsic and artificial (via the inhomogeneous magnetic field) spin-orbit interaction. We describe this effect by $\langle f_{\{Q\}_1} \rangle \rightarrow \langle f_{\{Q\}_1} \rangle + \delta f_{\{Q\}_1}$ and $\langle f_{\{Q\}_2} \rangle \rightarrow \langle f_{\{Q\}_2} \rangle + \delta f_{\{Q\}_2}$, where $\langle \delta f_{\{Q\}_1} \rangle$ and $\langle \delta f_{\{Q\}_2} \rangle$ are the single-qubit frequency fluctuations. Charge noise can additionally affect both qubits via correlated frequency shifts and the exchange interaction through the barrier voltage, which we model as $v_B \rightarrow v_B + \delta v_B$. In the presence of finite exchange coupling, one can define four distinct, pure dephasing times, each corresponding to the dephasing of a single qubit with the other qubit in a specific basis state. In a quasistatic approximation, the four dephasing times are then given by

$$\begin{aligned} & \$\$ \{T\}_2^{\{2\}} \langle \{Q\}_1 \rangle_1 (\langle \{Q\}_1 \rangle_2 = |0\rangle) = \frac{1}{\sqrt{2}} \frac{\langle \{Q\}_1 \rangle_1 \langle \{Q\}_2 \rangle_2 - \langle \{Q\}_1 \rangle_2 \langle \{Q\}_2 \rangle_1}{\sqrt{2}} \\ & \langle \{Q\}_1 \rangle_1 = \frac{1}{2} \left(\langle \{Q\}_1 \rangle_1 + \langle \{Q\}_1 \rangle_2 \right) + \frac{i}{2} \left(\langle \{Q\}_1 \rangle_1 - \langle \{Q\}_1 \rangle_2 \right) e^{i \omega_B t} \\ & \langle \{Q\}_2 \rangle_2 = \frac{1}{2} \left(\langle \{Q\}_1 \rangle_1 + \langle \{Q\}_1 \rangle_2 \right) - \frac{i}{2} \left(\langle \{Q\}_1 \rangle_1 - \langle \{Q\}_1 \rangle_2 \right) e^{i \omega_B t} \end{aligned} \quad (8)$$

$$\begin{aligned} & \$\$ \{T\}_2^{\{2\}} \langle \{Q\}_1 \rangle_1 (\langle \{Q\}_1 \rangle_2 = |1\rangle) = \frac{1}{\sqrt{2}} \frac{\langle \{Q\}_1 \rangle_1 \langle \{Q\}_2 \rangle_2 - \langle \{Q\}_1 \rangle_2 \langle \{Q\}_2 \rangle_1}{\sqrt{2}} \\ & \langle \{Q\}_1 \rangle_1 = \frac{1}{2} \left(\langle \{Q\}_1 \rangle_1 + \langle \{Q\}_1 \rangle_2 \right) + \frac{i}{2} \left(\langle \{Q\}_1 \rangle_1 - \langle \{Q\}_1 \rangle_2 \right) e^{i \omega_B t} \\ & \langle \{Q\}_2 \rangle_2 = \frac{1}{2} \left(\langle \{Q\}_1 \rangle_1 + \langle \{Q\}_1 \rangle_2 \right) - \frac{i}{2} \left(\langle \{Q\}_1 \rangle_1 - \langle \{Q\}_1 \rangle_2 \right) e^{i \omega_B t} \end{aligned} \quad (9)$$

$$\begin{aligned} & \$\$ \{T\}_2^{\{2\}} \langle \{Q\}_1 \rangle_2 (\langle \{Q\}_1 \rangle_1 = |0\rangle) = \frac{1}{\sqrt{2}} \frac{\langle \{Q\}_1 \rangle_1 \langle \{Q\}_2 \rangle_2 - \langle \{Q\}_1 \rangle_2 \langle \{Q\}_2 \rangle_1}{\sqrt{2}} \\ & \langle \{Q\}_1 \rangle_1 = \frac{1}{2} \left(\langle \{Q\}_1 \rangle_1 + \langle \{Q\}_1 \rangle_2 \right) + \frac{i}{2} \left(\langle \{Q\}_1 \rangle_1 - \langle \{Q\}_1 \rangle_2 \right) e^{i \omega_B t} \\ & \langle \{Q\}_2 \rangle_2 = \frac{1}{2} \left(\langle \{Q\}_1 \rangle_1 + \langle \{Q\}_1 \rangle_2 \right) - \frac{i}{2} \left(\langle \{Q\}_1 \rangle_1 - \langle \{Q\}_1 \rangle_2 \right) e^{i \omega_B t} \end{aligned} \quad (10)$$

$$\begin{aligned} & \$\$ \{T\}_2^{\{2\}} \langle \{Q\}_1 \rangle_2 (\langle \{Q\}_1 \rangle_1 = |1\rangle) = \frac{1}{\sqrt{2}} \frac{\langle \{Q\}_1 \rangle_1 \langle \{Q\}_2 \rangle_2 - \langle \{Q\}_1 \rangle_2 \langle \{Q\}_2 \rangle_1}{\sqrt{2}} \\ & \langle \{Q\}_1 \rangle_1 = \frac{1}{2} \left(\langle \{Q\}_1 \rangle_1 + \langle \{Q\}_1 \rangle_2 \right) + \frac{i}{2} \left(\langle \{Q\}_1 \rangle_1 - \langle \{Q\}_1 \rangle_2 \right) e^{i \omega_B t} \\ & \langle \{Q\}_2 \rangle_2 = \frac{1}{2} \left(\langle \{Q\}_1 \rangle_1 + \langle \{Q\}_1 \rangle_2 \right) - \frac{i}{2} \left(\langle \{Q\}_1 \rangle_1 - \langle \{Q\}_1 \rangle_2 \right) e^{i \omega_B t} \end{aligned}$$

Fitting qubit frequencies and dephasing times

The transition energies in equations (4)–(7) are fitted simultaneously to the measured results from the Ramsey experiment (see Fig. 3a). For the fitting, we use the NonLinearModelFit function from the software Mathematica with the least squares method. The best fits yield the following parameters: $\alpha = 12.1 \pm 0.05 \text{ V}^{-1}$, $\beta_1 = -2.91 \pm 0.11 \text{ MHz V}^{-\gamma}$, $\beta_2 = 67.2 \pm 0.63 \text{ MHz V}^{-\gamma}$, $\gamma = 1.20 \pm 0.01$ and $J_{\text{res}} = 58.8 \pm 1.8 \text{ kHz}$.

The dephasing times in equations (8)–(11) are fitted simultaneously to the measured results from the Ramsey experiment (see Fig. 3c) using the same method. The best fits yield the following parameters: $\delta\nu_B = 0.40 \pm 0.01$ mV, $\langle|\Delta f|_{\{Q\}_1}\rangle = 11 \pm 0.1$ kHz and $\langle|\Delta f|_{\{Q\}_2}\rangle = 24 \pm 0.7$ kHz).

Numerical simulations

For all numerical simulations, we solve the time-dependent Schrödinger equation

$$\langle \psi(t) | H | \psi(t) \rangle = \frac{d}{dt} \left(\frac{\partial}{\partial t} \langle \psi(t) | \hat{p} \psi(t) \rangle \right) - \frac{1}{2m} \frac{d^2}{dt^2} \langle \psi(t) | \hat{r}^2 \psi(t) \rangle$$

(12)

and iteratively compute the unitary propagator according to

\$\$U(t+\Delta t) = \{ e \}^{-i/\hbar} H(t+\Delta t) U(t) \quad (1)

where $\hbar = h/(2\pi)$ is the reduced Planck's constant. Here $H(t + \Delta t)$ is discretized into N segments of length Δt such that $H(t)$ is constant in the time interval $[t, t + \Delta t]$. All simulations are performed in the rotating frame of the external magnetic field $(B_{z,1} + B_{z,2})/2$ and neglecting the counter-rotating terms, making the so-called rotating-wave approximation. This allows us to choose $\Delta t = 10$ ps as a sufficiently small time step.

For the noise simulations, we included classical fluctuations of $\langle f_{Q_1} \rangle$ to $\langle f_{Q_1} \rangle + \delta f_{Q_1}$, $\langle f_{Q_2} \rangle$ to $\langle f_{Q_2} \rangle + \delta f_{Q_2}$ and $v_B \rightarrow v_B + \delta v_B$. We assume the noise coupling to the resonance

frequencies $\langle\delta f\rangle_{\{Q\}_1}\rangle$ and $\langle\delta f\rangle_{\{Q\}_2}\rangle$ to be quasistatic and assume $1/f$ noise for v_B , which we describe by its spectral density $\langle S_v \rangle_{\{\rm rm\{B\}\}}(\omega) = \delta(v - \{\rm rm\{B\}\})\Delta\omega$, where ω is the angular frequency. To compute time traces of the fluctuation, we use the approach introduced in refs. [45,46](#) to generate time-correlated time traces. The fluctuations are discretized into N segments with time Δt such that $\delta v_B(t)$ is constant in the time interval $[t, t + \Delta t]$, with the same Δt as above. Consequently, fluctuations that are faster than $\langle f \rangle_{\{\rm rm\{\max\}}}/\Delta t$ are truncated.

CZ gate

We realize a universal CZ = $\text{diag}(1, 1, 1, -1)$ gate by adiabatically pulsing the exchange interaction using a carefully designed pulse shape. Starting from equation [\(2\)](#), the full dynamics can be projected on the odd-parity space spanned by $|01\rangle$ and $|10\rangle$. The entangling exchange gate is reduced in this subspace to a global phase shift, thus, the goal is to minimize any dynamics inside the subspace. Introducing a new set of Pauli operators in this subspace $\sigma_x = |01\rangle\langle 10| + |10\rangle\langle 01|$, $\sigma_y = -i|01\rangle\langle 10| + i|10\rangle\langle 01|$ and $\sigma_z = |01\rangle\langle 01| - |10\rangle\langle 10|$, we find

$$\begin{aligned} \text{\$\$}\{H\}_{\{\rm rm\{sub\}\}}(t) &= \frac{1}{2}(-hJ(v)\{\rm rm\{B\}\}(t) + \Delta E_z) \sigma_z \\ &+ hJ(v)\{\rm rm\{B\}\}(t)\sigma_x. \end{aligned} \quad (14)$$

In order to investigate the adiabatic behaviour, it is convenient to switch into the adiabatic frame defined by $\langle U \rangle_{\{\rm rm\{ad\}\}} = \{\rm rm\{e\}\}^{\frac{1}{2}} \tan^{-1}(\frac{hJ(v)\{\rm rm\{B\}\}(t)}{\Delta E_z}) \sigma_y$. The Hamiltonian accordingly transforms as

$$\begin{aligned} \text{\$\$}\{H\}_{\{\rm rm\{ad\}\}} &= \{U\}_{\{\rm rm\{ad\}\}}^{\dagger} H_{\{\rm rm\{sub\}\}}(t) \\ &\{U\}_{\{\rm rm\{ad\}\}}(t) - \{\rm rm\{i\}\}\hbar \dot{U}_{\{\rm rm\{ad\}\}}^{\dagger} H_{\{\rm rm\{sub\}\}}(t) \\ &\{U\}_{\{\rm rm\{ad\}\}}(t) \end{aligned} \quad (15)$$

$$\begin{aligned} \text{\$\$}\approx \frac{1}{2}(-hJ(v)\{\rm rm\{B\}\}(t) + \varDelta E_z) \sigma_z - \\ \frac{1}{2}h^2 \dot{J}^2 \Delta E_z \sigma_y, \end{aligned} \quad (16)$$

where the first term is unaffected and describes the global phase accumulation due to the exchange interaction, the second term describes the single-qubit phase accumulations and the last term, $\langle f(t) = h^2 \dot{J}^2 / (4\pi) \Delta E_z \rangle$, describes the diabatic deviation proportional to the derivative of the exchange pulse. From equation [\(15\)](#) and equation [\(16\)](#), we assumed a constant $\Delta E_z(t)$

$\approx \Delta E_z$ and $hJ(t) \ll \Delta E_z$. The transition probability from state $|\uparrow\downarrow\rangle$ to $|\downarrow\uparrow\rangle$ using a pulse of length t_p is then given by³⁴

$$\begin{aligned} & \$\$ \{P\}_{\{\uparrow\downarrow\}} = \int_0^{\infty} dt f(t) e^{-iE_z t / \hbar} \Delta E \\ & \text{(17)} \end{aligned}$$

$$\$\$ \propto S_s(f(t)). \$\$ \quad (18)$$

From the first to the second line, we identify the integral by the (short-timescale) Fourier transform, allowing us to describe the spin-flip error probability by the energy spectral density S_s of the input signal $f(t)$. Minimizing such errors is, therefore, identical to minimizing the energy spectral density of a pulse, a well-known and solved problem from classical signal processing and statistics. Optimal shapes are commonly referred to as window functions $W(t)$ due to their property of restricting the spectral resolution of signals. A high-fidelity exchange pulse is consequently given by $J(0)=J(t_p)$ and

$$\begin{aligned} & \$\$ \int_0^{\infty} dt J(t) W(t) = 1/4, \$\$ \\ & \text{(19)} \end{aligned}$$

while setting $J(t) = A_{\rm v} W(t) / J_{\rm res}$ (ref. 34), with a scaling factor $A_{\rm v}$ that is to be determined. In this work, we have chosen the cosine window

$$\begin{aligned} & \$\$ W(t) = \frac{1}{2} \left[1 - \cos \left(\frac{2\pi}{t_p} t \right) \right] \\ & \text{(20)} \end{aligned}$$

from signal processing, which has a high spectral resolution. The amplitude $A_{\rm v}$ follows from condition equation (19). For a pulse length of $t_p = 100$ ns and a cosine pulse shape, we find $A_{\rm v} = 10.06 \text{ MHz}$. As explained in the main text, owing to the exponential voltage-exchange relation, the target pulse shape for $J(t)$ must be converted to a barrier gate pulse, following⁴⁷

$$\begin{aligned} & \$\$ v_{\rm B}(t) = \frac{1}{2} \alpha \log \left(\frac{A_{\rm v}}{W(t)} \right), \$\$ \\ & \text{(21)} \end{aligned}$$

Our numerical simulations predict an average gate infidelity $1 - F_{\text{gate}} < 10^{-6}$ without noise and $1 - F = 0.22 \times 10^{-3}$ with the inclusion of noise through the fluctuations

$(\delta f_{Q_1}), (\delta f_{Q_2})$ and δv_B , discussed in the previous section. The measured PTMs reveal much higher rates of incoherent errors, which we attribute to drifts in the barrier voltage on a timescale much longer than the timescale on which $(\delta f_{Q_1}), (\delta f_{Q_2})$ and δv_B were determined.

Gate-set tomography analysis

We designed a GST experiment using the gate set $\{\{I\}\}, \{\{X\}\}_{\{Q\}}_{\{1\}}, \{\{Y\}\}_{\{Q\}}_{\{1\}}, \{\{X\}\}_{\{Q\}}_{\{2\}}, \{\{Y\}\}_{\{Q\}}_{\{2\}}, \{\{CZ\}\}$, where I is a 100-ns idle gate, $\{\{X\}\}_{\{Q\}}_{\{1\}}$ and $\{\{Y\}\}_{\{Q\}}_{\{1\}}$ are single-qubit $\pi/2$ gates with rotation axis (\hat{x}) ((\hat{y})) on Q_1 and Q_2 , with durations of 150 ns and 200 ns, respectively, and $CZ = \text{diag}(1, 1, 1, -1)$. A classic two-qubit GST experiment consists of a set of germs designed to amplify all types of error in the gate set when repeated and a set of 36 fiducials composed by the 11 elementary operations $\{\{null\}\}, \{\{X\}\}_{\{Q\}}_{\{1\}}, \{\{Y\}\}_{\{Q\}}_{\{1\}}, \{\{X\}\}_{\{Q\}}_{\{2\}}, \{\{Y\}\}_{\{Q\}}_{\{2\}}$

$\{\{X\}\}_{\{Q\}}_{\{1\}}, \{\{Y\}\}_{\{Q\}}_{\{1\}}, \{\{X\}\}_{\{Q\}}_{\{2\}}, \{\{Y\}\}_{\{Q\}}_{\{2\}}$ required to carry out quantum process tomography of the germs⁴⁸. We use a set of 16 germs $\{\{I\}, \{X\}_{\{Q\}}_{\{1\}}, \{Y\}_{\{Q\}}_{\{1\}}, \{X\}_{\{Q\}}_{\{2\}}, \{Y\}_{\{Q\}}_{\{2\}}, \{CZ\}, \{X\}_{\{Q\}}_{\{1\}}, \{Y\}_{\{Q\}}_{\{1\}}, \{X\}_{\{Q\}}_{\{2\}}, \{Y\}_{\{Q\}}_{\{2\}}, \{X\}_{\{Q\}}_{\{1\}}, \{Y\}_{\{Q\}}_{\{1\}}, \{X\}_{\{Q\}}_{\{2\}}, \{Y\}_{\{Q\}}_{\{2\}}, \{X\}_{\{Q\}}_{\{1\}}, \{Y\}_{\{Q\}}_{\{1\}}, \{X\}_{\{Q\}}_{\{2\}}, \{Y\}_{\{Q\}}_{\{2\}}\}$

$\{\{I\}, \{X\}_{\{Q\}}_{\{1\}}, \{Y\}_{\{Q\}}_{\{1\}}, \{X\}_{\{Q\}}_{\{2\}}, \{Y\}_{\{Q\}}_{\{2\}}, \{CZ\}, \{X\}_{\{Q\}}_{\{1\}}, \{Y\}_{\{Q\}}_{\{1\}}, \{X\}_{\{Q\}}_{\{2\}}, \{Y\}_{\{Q\}}_{\{2\}}, \{X\}_{\{Q\}}_{\{1\}}, \{Y\}_{\{Q\}}_{\{1\}}, \{X\}_{\{Q\}}_{\{2\}}, \{Y\}_{\{Q\}}_{\{2\}}, \{X\}_{\{Q\}}_{\{1\}}, \{Y\}_{\{Q\}}_{\{1\}}, \{X\}_{\{Q\}}_{\{2\}}, \{Y\}_{\{Q\}}_{\{2\}}\}$

$\{\{\rm{rm\{X\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{1\}\}\{\{\rm{rm\{Y\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{2\}\}$
 $\{\{\rm{rm\{X\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{2\}\}\{\{\rm{rm\{Y\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{1\}\}$
 $\{\{\rm{rm\{X\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{2\}\}\{\{\rm{rm\{X\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{2\}\},\backslash(\{\{\rm{rm\{C\}}}\}$
 $\{\rm{rm\{Z\}}\}\{\rm{rm\{X\}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{2\}\}\{\{\rm{rm\{Y\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{1\}\}\{\{\rm{rm\{C\}}}\}$
 $\{\rm{rm\{Z\}}\}\{\rm{rm\{Y\}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{2\}\}\{\{\rm{rm\{X\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{1\}\},$
 $\{\{\rm{rm\{Y\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{1\}\}\{\{\rm{rm\{X\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{1\}\}$
 $\{\{\rm{rm\{Y\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{2\}\}\{\{\rm{rm\{X\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{1\}\}$
 $\{\{\rm{rm\{X\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{2\}\}\{\{\rm{rm\{X\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{1\}\}$
 $\{\{\rm{rm\{Y\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{1\}\}\{\{\rm{rm\{Y\}}}\}_{-}\{\{\{\rm{rm\{Q\}}}\}_{-}\{2\}\}\}\backslash)(\text{ref. } \textcolor{blue}{35})$. Note that the null gate is the instruction for doing nothing in zero time, different from the idle gate. Simple errors such as errors in the rotation angle of a particular gate can be amplified by simply repeating the same gate. More complicated errors such as tilts in rotation axes can only be amplified by a combination of different gates. The germs and fiducials are then compiled into GST sequences, such that each sequence consists of two fiducials interleaved by a single germ or power of germs³⁵ (as illustrated in Fig. 2a). The GST sequences are classified by their germ powers into lengths $L = 1, 2, 4, 8, 16\dots$, where a sequence of length n consists of n gates plus the fiducial gates. We note that the sequences used in GST are shorter than the sequences involved in other methods to self-consistently estimate the gate performance, such as randomized benchmarking. As a result, GST suffers less from drift in qubit frequencies and readout windows induced by long sequences of microwave bursts.

After the execution of all sequences, a maximum-likelihood estimation is performed to estimate the process matrices of each gate in the gate set and the SPAM probabilities. We use the open source pyGSTi Python package^{49,50} to perform the maximum-likelihood estimation, as well as to design an optimized GST experiment by eliminating redundant circuits and to provide statistical error bars by computing all involved Hessians. The circuit optimization allows us to perform GST with a maximum sequence length $L_{\max} = 16$ using 1,685 different sequences in total. The pyGSTi package quantifies the Markovian-model violation of the experimental data, counting the number of standard deviations exceeding their expectation values under the χ^2 hypothesis⁵⁰. This model violation is internally translated into a more accessible goodness ratio from 0 to 5, with 5 being the best⁴⁹, where we obtain a 4 out of 5 rating, indicating remarkably small deviations from expected results. The total number of standard deviations exceeding the expected results for each L , as well as the contribution of each sequence to this number, can be found in the pyGSTi report, along with the supporting data.

From the GST experiment, we have extracted the PTM $\{\{\mathcal{M}\}\}_{-}\{\exp\}$ describing each gate in our gate set $\{\{\{\rm{rm\{I\}}\},\{\{\rm{rm\{X\}}\}\}_{-}\{\{\{\rm{rm\{Q\}}\}\}_{-}\{1\}\},$
 $\{\{\rm{rm\{Y\}}\}\}_{-}\{\{\{\rm{rm\{Q\}}\}\}_{-}\{1\}\},\{\{\rm{rm\{X\}}\}\}_{-}\{\{\{\rm{rm\{Q\}}\}\}_{-}\{2\}\},$
 $\{\{\rm{rm\{Y\}}\}\}_{-}\{\{\{\rm{rm\{Q\}}\}\}_{-}\{2\}\},\{\rm{rm\{CZ\}}\}\}$. The PTM is isomorphically related

to the conventionally used χ matrix describing a quantum process. A completely positive, trace-preserving, two-qubit PTM has 240 parameters describing the process. To obtain insight into the errors of the gates in the experiment, we first compute the error in the PTM given by $\langle E = \{ \mathcal{M} \} \rangle_{\text{exp}} \{ \mathcal{M} \}_{\{\text{ideal}\}}^{\{-1\}}$, where we have adapted the convention to add the error after the ideal gate. The average gate fidelity is then conveniently given by

$$\text{F} = \frac{\text{Tr}(\mathcal{M})}{\text{Tr}(\mathcal{M}_{\{\text{ideal}\}})} \quad (22)$$

It is related to the entanglement fidelity via $\langle 1 - F \rangle_{\{\text{ent}\}} = \frac{d+1}{d}(1 - \langle F \rangle_{\{\text{gate}\}})$ (ref. [51](#)), where d is the dimension of the two-qubit Hilbert space. Although the PTM $\langle \mathcal{M} \rangle$ perfectly describes the errors, it is more intuitive to analyse the corresponding error generator $\langle \mathcal{L} \rangle = \log(E)$ of the process^{[30](#)}. The error generator $\langle \mathcal{L} \rangle$ relates to the error PTM E in a similar way as a Hamiltonian H relates to a unitary operation $U = e^{-iH}$. The error generator can be separated into several blocks. A full discussion about the error generator can be found in ref. [30](#). In this work, we have used the error generator to distinguish the dynamics originating from coherent Hamiltonian errors, which can be corrected by adjusting gate parameters (see Extended Data Fig. [5](#)), and from noisy/stochastic dynamics, which cannot be corrected easily. The coherent errors can be extracted by projecting $\langle \mathcal{L} \rangle$ onto the 4×4 -dimensional Hamiltonian space H . In the Hilbert–Schmidt space, the Hamiltonian projection is given by^{[30](#)}

$$\text{H}_{mn} = -\frac{1}{d^2} \text{Tr}[(P_m)^T \otimes P_n] \langle \mathcal{L} \rangle [P_m \otimes P_n], \quad (23)$$

where $\langle \mathcal{L} \rangle$ is the error generator in Liouville superoperator form, $P_m \in \{I, X, Y, Z\}$ are the extended Pauli matrices with $m, n = 0, 1, 2, 3$, $\mathbf{1}_d$ is the d -dimensional identity matrix and $d = 4$ is the dimension of the two-qubit Hilbert space. To improve the calibration of our gate set, we use the information of seven different Hamiltonian errors (IX, IY, XI, YI, ZI, IZ and ZZ). To estimate coherent Hamiltonian errors and incoherent stochastic errors, two new metrics are considered^{[30](#)}: the Jamiołkowski probability

$$\langle \epsilon \rangle = \langle \text{Tr}(\rho \langle \mathcal{L} \rangle | \Psi \rangle \langle \Psi |) \rangle \quad (24)$$

which describes the amount of incoherent error in the process, and the Jamiołkowski amplitude

$$\begin{aligned} \text{\$}\$\{\theta\}_{\rm J}(\mathcal{L}) &= \sqrt{(1 - |\Psi\rangle\langle\Psi|)\rho} \\ &\quad \times (\rho - |\Psi\rangle\langle\Psi|)^2, \end{aligned} \quad (25)$$

which approximately describes the amount of coherent Hamiltonian errors (Extended Data Table 1). Here $(\rho_{\rm J}(\mathcal{L}))^2 = |\Psi\rangle\langle\Psi|$ is the Jamiołkowski state and $|\Psi\rangle$ is a maximally entangling four-qubit state that originates from the relation of quantum processes to states in a Hilbert space twice the dimension via the Choi–Jamiołkowski isomorphism⁵². For small errors, the average gate infidelity can be approximated by³⁰

$$\begin{aligned} \text{\$}\$1 - F_{\rm gate} &= \frac{d}{d+1} [\epsilon_{\rm J}(\mathcal{L}) + \\ &\quad \theta_{\rm J}(\mathcal{L})^2]. \end{aligned} \quad (26)$$

For a comparison of the performance of the single-qubit gates with previous experiments reporting single-qubit gate fidelities, we compute the fidelities projected to the single-qubit space from the PTMs or the error generators. In Fig. 2 and Extended Data Fig. 2, single-qubit gate fidelities are estimated by projecting the PTMs onto the corresponding subspace. Let \mathcal{P}_j be the projector on the subspace of qubit j , then the fidelity is given by

$$\begin{aligned} F_{\rm sub} &= \frac{1}{d} \operatorname{Tr}(\mathcal{P}_j \mathcal{M} \exp(-\mathcal{M}^T \mathcal{M} / 2)) \\ &\quad + (d/2) \operatorname{Tr}(\mathcal{P}_j \mathcal{M} \exp(-\mathcal{M}^T \mathcal{M} / 2)). \end{aligned} \quad (27)$$

Error bars for the fidelity projected to the subspace are computed using standard error propagation of the confidence intervals of $\operatorname{Tr}(\mathcal{M} \exp(-\mathcal{M}^T \mathcal{M} / 2))$ provided by the pyGSTi package. A more optimistic estimation for the fidelities in the single-qubit subspace is given by projecting the error generators instead of the PTMs.

Variational quantum eigensolver

We follow the approach of ref. 36 to using the VQE algorithm to compute the ground-state energy of molecular hydrogen, after mapping this state onto the state of two qubits. We include this information here for completeness. The Hamiltonian of a molecular system in atomic units is

$$\begin{aligned}
 & \sum_{i,j} \left[\frac{\nabla}{\nabla} \cdot \left(\frac{\nabla}{\nabla} R_i \right) R_j - \frac{\nabla}{\nabla} \cdot \left(\frac{\nabla}{\nabla} R_j \right) R_i \right] \\
 & + \sum_{i,j} \left[\frac{\nabla}{\nabla} \cdot \left(\frac{\nabla}{\nabla} R_i \right) R_j - \frac{\nabla}{\nabla} \cdot \left(\frac{\nabla}{\nabla} R_j \right) R_i \right] \\
 & + \sum_{i,j} \left[\frac{\nabla}{\nabla} \cdot \left(\frac{\nabla}{\nabla} R_i \right) R_j - \frac{\nabla}{\nabla} \cdot \left(\frac{\nabla}{\nabla} R_j \right) R_i \right]
 \end{aligned}$$

where $\langle\langle\langle\bf{R}\rangle\rangle\rangle_i$, M_i and Q_i are the position, mass and charge, respectively, of the i th nuclei and $\langle\langle\langle\bf{r}\rangle\rangle\rangle_j$ is the position of the j th electron. The first two sums describe the kinetic energies of the nuclei and electrons, respectively. The last three sums describe the Coulomb repulsion between nuclei and electrons, nuclei and nuclei, and electrons and electrons, respectively. As we are primarily interested in the electronic structure of the molecule, and nuclear masses are a few orders of magnitude larger than the electron masses, the nuclei are treated as static point charges under the Born–Oppenheimer approximation. Consequently, the electronic Hamiltonian can be simplified to

$$\begin{aligned} & \sum_{i,j} (\nabla \cdot \mathbf{r}_i) = -\sum_i \frac{\partial}{\partial r_i} \left(\frac{Q_i}{r_i} - \frac{Q_j}{r_j} \right) + \sum_{i>j} \frac{1}{r_i - r_j}. \end{aligned} \quad (29)$$

Switching into the second-quantization representation, described by fermionic creation and annihilation operators, $\langle\{a\}_p\rangle^\dagger$ and a_q , acting on a finite basis, the Hamiltonian becomes

$$\begin{aligned} & \sum_{\{H\}} \{\{ \text{rm}\{e\} \}\} = \sum_{\{pq\}} \{h\}_{\{pq\}} \{a\}_{\{p\}}^{\dagger} \{a\}_{\{q\}} + \sum_{\{pqrs\}} \\ & \{h\}_{\{pqrs\}} \{a\}_{\{p\}}^{\dagger} \{a\}_{\{q\}}^{\dagger} \{a\}_{\{r\}} \{a\}_{\{s\}}, \end{aligned} \quad (30)$$

where p, q, r and s label the corresponding basis states. The antisymmetry under exchange is retained through the anticommutation relation of the operators. The weights of the two sums are given by the integrals

$$\begin{aligned} \text{\$}\$\{h\}_{pq}=&\int \{\rm d\} \{\boldsymbol{\sigma}\} \{\psi\} \{p\}^{\ast} \\ &(\{\boldsymbol{\sigma}\})(\frac{\nabla}{\{r\}_i^2})^2 - \sum_i \frac{Q_i}{\{|R_i - r|\}} \psi_q(\boldsymbol{\sigma}), \end{aligned}$$

```
 $$\langle h \rangle_{pqrs} = \int \{ \mathrm{d} \} \{ \boldsymbol{\sigma} \} \{ \boldsymbol{\sigma} \} \frac{ \{ \psi \}^p \{ \psi \}^q }{ \{ \psi \}^r \{ \psi \}^s } \langle \{ \boldsymbol{\sigma} \} \{ \boldsymbol{\sigma} \} \rangle
```

$$\} \}_{\{2\}} \{ \{ \mathrm{rm}\{\psi\} \} \}_{\{s\}} (\{ \{ \boldsymbol{\sigma} \} \}_{\{1\}} \{ \{ \mathrm{rm}\{\psi\} \} \}_{\{r\}} \\ (\{ \{ \boldsymbol{\sigma} \} \}_{\{2\}}) \{ \{ \mathrm{bf}\{r\} \} \}_{\{1\}} - \{ \{ \mathrm{bf}\{r\} \} \}_{\{2\}} \}, \$\$ \\ (32)$$

where $\{ \{ \boldsymbol{\sigma} \} \}_{\{i\}}$ ($\{ \{ \mathrm{bf}\{r\} \} \}_{\{i\}}, \{s\}_{\{i\}}$) is a multi-index describing the position $\{ \{ \mathrm{bf}\{r\} \} \}_{\{i\}}$ and the spin s_i of electron i . Such a second-quantized molecular Hamiltonian can be mapped onto qubits using the Jordan–Wigner (JW) or the BK transformation⁵. The JW transformation directly encodes the occupation number (0 or 1) of the i th spin orbital into the state ($|0\rangle$ or $|1\rangle$) of the i th qubit. The number of qubits required after JW transformation is, thus, the same as the number of spin orbitals that are of interest. The BK transformation, on the other hand, encodes the information in both the occupation number and parities, whether there is an even or odd occupation in a subset of spin orbitals.

Taking molecular hydrogen in the HF basis as an example, we are interested in investigating the bonding ($|O_1\uparrow\rangle, |O_1\downarrow\rangle$) and the antibonding orbital state ($|O_2\uparrow\rangle, |O_2\downarrow\rangle$). The initial guess of the solution is the HF state in which both electrons occupy the $|O_1\rangle$ orbital. The JW transformation encodes the HF initial state as $|1100\rangle$, representing $\langle (|N\rangle_{\{O\}_{\{1\}}} \downarrow \rangle \{N\}_{\{O\}_{\{1\}}} \uparrow \rangle \{N\}_{\{O\}_{\{2\}}} \downarrow \rangle \{N\}_{\{O\}_{\{2\}}} \uparrow \rangle \rangle$ from left to right, where $\langle (|N\rangle_{\{O\}_{\{i\}}} S \rangle)$ is the occupation of the $O_i S$ spin orbital with $S = \uparrow, \downarrow$. The BK transformation encodes the HF initial state as $|1000\rangle$, where the first and the third qubits (counting from the right) encode the occupation number of the first and third spin orbitals ($\langle (|N\rangle_{\{O\}_{\{1\}}} \uparrow \rangle = 1 \rangle$ and $\langle (|N\rangle_{\{O\}_{\{2\}}} \uparrow \rangle = 0 \rangle$), the second qubit encodes the parity of the first two spin orbitals ($\langle (|N\rangle_{\{O\}_{\{1\}}} \uparrow \rangle + |N\rangle_{\{O\}_{\{1\}}} \downarrow \rangle) \bmod 2 = 0$) and the fourth qubit encodes the parity of all four spin orbitals ($\langle (|N\rangle_{\{O\}_{\{1\}}} \uparrow \rangle + |N\rangle_{\{O\}_{\{1\}}} \downarrow \rangle + |N\rangle_{\{O\}_{\{2\}}} \uparrow \rangle + |N\rangle_{\{O\}_{\{2\}}} \downarrow \rangle) \bmod 2 = 0$). With the standard transformation rules for fermionic creation and annihilation operators, the system Hamiltonian becomes a four-qubit Hamiltonian

$$\$\$ \begin{array}{c} \{H\}_{\{\mathrm{rm}\{JW\}\}} = \{g\}_{\{0\}} \{ \mathrm{rm}\{I\} \} + \{g\}_{\{1\}} \\ \{ \mathrm{rm}\{Z\} \}_{\{1\}} + \{g\}_{\{2\}} \{ \mathrm{rm}\{Z\} \}_{\{2\}} + \{g\}_{\{3\}} \{ \mathrm{rm}\{Z\} \}_{\{3\}} + \{g\}_{\{4\}} \\ \{ \mathrm{rm}\{Z\} \}_{\{4\}} \\ \{ \mathrm{rm}\{Z\} \}_{\{5\}} \{ \mathrm{rm}\{Z\} \}_{\{1\}} \{ \mathrm{rm}\{Z\} \}_{\{2\}} + \{g\}_{\{6\}} \\ \{ \mathrm{rm}\{Z\} \}_{\{1\}} \{ \mathrm{rm}\{Z\} \}_{\{3\}} + \{g\}_{\{7\}} \{ \mathrm{rm}\{Z\} \}_{\{1\}} \{ \mathrm{rm}\{Z\} \}_{\{4\}} \\ \{ \mathrm{rm}\{Z\} \}_{\{8\}} \{ \mathrm{rm}\{Z\} \}_{\{2\}} \{ \mathrm{rm}\{Z\} \}_{\{3\}} + \{g\}_{\{9\}} \{ \mathrm{rm}\{Z\} \}_{\{2\}} \\ \{ \mathrm{rm}\{Z\} \}_{\{4\}} + \{g\}_{\{10\}} \{ \mathrm{rm}\{Z\} \}_{\{3\}} \{ \mathrm{rm}\{Z\} \}_{\{4\}} \\ \{ \mathrm{rm}\{Y\} \}_{\{1\}} \{ \mathrm{rm}\{X\} \}_{\{2\}} \{ \mathrm{rm}\{X\} \}_{\{3\}} \{ \mathrm{rm}\{Y\} \}_{\{4\}} + \{g\}_{\{11\}} \\ \{ \mathrm{rm}\{Y\} \}_{\{1\}} \{ \mathrm{rm}\{Y\} \}_{\{2\}} \{ \mathrm{rm}\{X\} \}_{\{3\}} \{ \mathrm{rm}\{X\} \}_{\{4\}} + \{g\}_{\{12\}} \\ \{ \mathrm{rm}\{Y\} \}_{\{1\}} \{ \mathrm{rm}\{Y\} \}_{\{2\}} \{ \mathrm{rm}\{X\} \}_{\{3\}} \{ \mathrm{rm}\{X\} \}_{\{4\}} \\ \{ \mathrm{rm}\{X\} \}_{\{1\}} \{ \mathrm{rm}\{X\} \}_{\{2\}} \{ \mathrm{rm}\{Y\} \}_{\{3\}} \{ \mathrm{rm}\{Y\} \}_{\{4\}} + \{g\}_{\{13\}} \\ \{ \mathrm{rm}\{X\} \}_{\{1\}} \{ \mathrm{rm}\{Y\} \}_{\{2\}} \{ \mathrm{rm}\{Y\} \}_{\{3\}} \{ \mathrm{rm}\{X\} \}_{\{4\}} + \{g\}_{\{14\}} \\ \{ \mathrm{rm}\{X\} \}_{\{1\}} \{ \mathrm{rm}\{Y\} \}_{\{2\}} \{ \mathrm{rm}\{Y\} \}_{\{3\}} \{ \mathrm{rm}\{X\} \}_{\{4\}} \end{array} \$\$$$

(33)

$$\begin{aligned}
 & \$\$ \begin{array}{l} \{H\}_{\{\{\rm BK\}\}} = \{g\}_0 \{\rm I\} + \{g\}_1 \\
 & \{\{\rm Z\}\}_{\{1\}} + \{g\}_2 \{\{\rm Z\}\}_{\{2\}} + \{g\}_3 \{\{\rm Z\}\}_{\{3\}} \\
 & \{\{\rm Z\}\}_{\{1\}} \{\{\rm Z\}\}_{\{2\}} + \{g\}_5 \{\{\rm Z\}\}_{\{1\}} \{\{\rm Z\}\}_{\{3\}} + \\
 & \{g\}_6 \{\{\rm Z\}\}_{\{2\}} \{\{\rm Z\}\}_{\{4\}} \\
 & \{\{\rm Z\}\}_{\{2\}} \{\{\rm Z\}\}_{\{3\}} + \{g\}_8 \{\{\rm Z\}\}_{\{1\}} \{\{\rm Z\}\}_{\{3\}} \\
 & \{\{\rm Z\}\}_{\{4\}} + \{g\}_9 \{\{\rm Z\}\}_{\{2\}} \{\{\rm Z\}\}_{\{3\}} \{\{\rm Z\}\}_{\{4\}} \\
 & \{g\}_{10} \{\{\rm Z\}\}_{\{1\}} \{\{\rm Z\}\}_{\{2\}} \{\{\rm Z\}\}_{\{3\}} \{\{\rm Z\}\}_{\{4\}} + \\
 & \{g\}_{11} \{\{\rm X\}\}_{\{1\}} \{\{\rm Z\}\}_{\{2\}} \{\{\rm X\}\}_{\{3\}} \\
 & \{\{\rm Y\}\}_{\{1\}} \{\{\rm Z\}\}_{\{2\}} \{\{\rm Y\}\}_{\{3\}} + \{g\}_{13} \{\{\rm X\}\}_{\{1\}} \\
 & \{\{\rm Z\}\}_{\{2\}} \{\{\rm X\}\}_{\{3\}} \{\{\rm Z\}\}_{\{4\}} + \{g\}_{14} \{\{\rm Y\}\}_{\{1\}} \\
 & \{\{\rm Z\}\}_{\{2\}} \{\{\rm Y\}\}_{\{3\}} \{\{\rm Z\}\}_{\{4\}} \end{array} \$\$ \\
 \end{aligned}$$

(34)

The subscripts are used to label the qubits. We see that, owing to the symmetry of the represented system in H_{BK} , qubit 2 and qubit 4 are never flipped, allowing us to reduce the dimension of the Hamiltonian to

$$\begin{aligned}
 & \$\$ \begin{array}{l} \{c\}_{\{\{\rm BK\}\}}^{\{\{\rm reduced\}\}} = \{h\}_0 \{\rm I\} + \\
 & \{h\}_1 \{\{\rm Z\}\}_{\{1\}} + \{h\}_2 \{\{\rm Z\}\}_{\{2\}} + \{h\}_3 \{\{\rm Z\}\}_{\{1\}} \\
 & \{\{\rm Z\}\}_{\{2\}} \\
 & \{h\}_4 \{\{\rm X\}\}_{\{1\}} \{\{\rm X\}\}_{\{2\}} + \{h\}_5 \\
 & \{\{\rm Y\}\}_{\{1\}} \{\{\rm Y\}\}_{\{2\}} \{h\}_0 \\
 & \{\rm IZ\} + \{h\}_3 \{\rm ZZ\} + \{h\}_4 \{\rm XX\} + \{h\}_5 \\
 & \{\rm YY\}, \end{array} \$\$ \\
 \end{aligned}$$

(35)

where qubit 1 has been relabelled as qubit 2 and qubit 3 has been relabelled as qubit 1. The HF initial state is, therefore, reduced to $|01\rangle$ and the Hamiltonian is rephrased to be consistent with the partial tomography expression in Fig. 5. This reduced representation requires only two qubits to simulate the hydrogen molecule. We emphasize that such a reduction of the BK Hamiltonian is not a special case for the H_2 molecule but is connected to symmetry considerations to reduce the complexity of systems, in a scalable way.

VQE is a method to compute the ground-state energy of the Hamiltonian. The total energy can be directly calculated by measuring the expectation value of each Hamiltonian term. This can be done easily by partial quantum state tomography. All the expectation values are then added up with a set of weights (h_0 through h_5). The weights are only functions of the internuclear separation (R) and can be computed efficiently by a classical computer. Here we use the OpenFermion Python package to compute these weights³⁸.

The main task of the quantum processor is, then, to encode the molecular spin-orbital state into the qubits. The starting point is the HF initial state, which is believed to largely overlap with the actual ground state. In order to find the actual ground state, the initial state needs to be ‘parameterized’ into an ansatz to explore a subspace of all possible states. We apply the unitary coupled cluster (UCC) theory to the parameterized ansatz state, which is used to describe many-body systems and cannot be efficiently executed on a classical computer⁵³. The UCC operator has a format

$$\$ \$ \{U\}_\{\{\rm{UCC}\}\}(\{\boldsymbol{\theta}\}) = \{\{\rm{e}\}\}^\wedge \{\sum_n (T_n (\{\boldsymbol{\theta}\}) - T_n^\dagger (\{\boldsymbol{\theta}\}))\}, \$ \$$$

(36)

with

$$\$ \$ \{T\}_1 (\{\boldsymbol{\theta}\}) = \sum_{m,i} \{\boldsymbol{\theta}\}_{i,m}^m, \$ \$$$

(37)

$$\$ \$ \{T\}_2 (\{\boldsymbol{\theta}\}) = \sum_{m,n,i,j} \{\boldsymbol{\theta}\}_{i,j}^{m,n} \{a\}_{i,m}^m \{a\}_{j,n}^n, \$ \$$$

(38)

representing single and double excitation of the electrons. The indices i and j label the occupied spin orbitals and m and n are the labels of the unoccupied spin orbitals. The vector θ is the set of all parameters to optimize. In the case of a H₂ molecule, the UCC operator is transformed into a qubit operator as

$$\$ \$ \{U\}_\{\{\rm{UCC}\}\}^\wedge (\{\rm{BK}\}) (\{\boldsymbol{\theta}\}) = \{\{\rm{e}\}\}^\wedge (-\{\rm{i}\}\theta \{\rm{XY}\}), \$ \$$$

where θ is a single parameter to variationally optimize.

Data availability

Data supporting this work are available at Zenodo,
<https://doi.org/10.5281/zenodo.5044450>.

Code availability

The codes used for data acquisition and processing are from the open source Python packages QCoDeS (<https://github.com/QCoDeS/Qcodes>), QTT (<https://github.com/QuTech-Delft/qtt>) and PycQED (https://github.com/DiCarloLab-Delft/PycQED_py3). The codes used for the design and analysis of the gate-set

tomography experiment are from pyGSTi (<https://github.com/pyGSTi/pyGSTi>). The codes used for the design and analysis of the variational quantum eigensolver experiment are from OpenFermion (<https://github.com/quantumlib/OpenFermion>).

References

1. 1.

Lidar, D. A. & Brun, T. A. *Quantum Error Correction* (Cambridge Univ. Press, 2013).

2. 2.

Raussendorf, R. & Harrington, J. Fault-tolerant quantum computation with high threshold in two dimensions. *Phys. Rev. Lett.* **98**, 190504 (2007).

3. 3.

Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).

4. 4.

Zwerver, A. M. J. et al. Qubits made by advanced semiconductor manufacturing. Preprint at <https://arxiv.org/abs/2101.12650> (2021).

5. 5.

McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S. C. & Yuan, X. Quantum computational chemistry. *Rev. Mod. Phys.* **92**, 015003 (2020).

6. 6.

Nielsen, M. A. & Chuang, I. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2002).

7. 7.

Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).

8. 8.

Veldhorst, M. et al. A two-qubit logic gate in silicon. *Nature* **526**, 410–414 (2015).

9. 9.

Zajac, D. M., Hazard, T. M., Mi, X., Nielsen, E. & Petta, J. R. Scalable gate architecture for a one-dimensional array of semiconductor spin qubits. *Phys. Rev. Appl.* **6**, 054013 (2016).

10. 10.

Vandersypen, L. M. K. et al. Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent. *npj Quantum Inf.* **3**, 34 (2017).

11. 11.

Yoneda, J. et al. A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%. *Nat. Nanotechnol.* **13**, 102–106 (2018).

12. 12.

Yang, C. H. et al. Silicon qubit fidelities approaching incoherent noise limits via pulse engineering. *Nat. Electron.* **2**, 151–158 (2019).

13. 13.

Xue, X. et al. Benchmarking gate fidelities in a Si/SiGe two-qubit device. *Phys. Rev. X* **9**, 021011 (2019).

14. 14.

Huang, W. et al. Fidelity benchmarks for two-qubit gates in silicon. *Nature* **569**, 532–536 (2019).

15. 15.

Takeda, K. et al. Quantum tomography of an entangled three-qubit state in silicon. *Nat. Nanotechnol.* **16**, 965–969 (2021).

16. 16.

Watson, T. F. et al. A programmable two-qubit quantum processor in silicon. *Nature* **555**, 633–637 (2018).

17. 17.

Xue, X. et al. CMOS-based cryogenic control of silicon quantum circuits. *Nature* **593**, 205–210 (2021).

18. 18.

Elzerman, J. M. et al. Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430**, 431–435 (2004).

19. 19.

Pioro-Ladrière, M. et al. Electrically driven single-electron spin resonance in a slanting Zeeman field. *Nat. Phys.* **4**, 776–779 (2008).

20. 20.

Vandersypen, L. M. K. & Chuang, I. L. NMR techniques for quantum control and computation. *Rev. Mod. Phys.* **76**, 1037 (2005).

21. 21.

Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120 (1998).

22. 22.

Petta, J. R. et al. Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005).

23. 23.

Meunier, T., Calado, V. E. & Vandersypen, L. M. K. Efficient controlled-phase gate for single-spin qubits in quantum dots. *Phys. Rev. B* **83**, 121403 (2011).

24. 24.

Martins, F. et al. Noise suppression using symmetric exchange gates in spin qubits. *Phys. Rev. Lett.* **116**, 116801 (2016).

25. 25.

Reed, M. D. et al. Reduced sensitivity to charge noise in semiconductor spin qubits via symmetric operation. *Phys. Rev. Lett.* **116**, 110402 (2016).

26. 26.

Blume-Kohout, R. et al. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nat. Commun.* **8**, 14485 (2017).

27. 27.

Magesan, E., Gambetta, J. M. & Emerson, J. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A* **85**, 042311 (2012).

28. 28.

Dehollain, J. P. et al. Optimization of a solid-state electron spin qubit using gate set tomography. *New J. Phys.* **18**, 103018 (2016).

29. 29.

White, G. A., Hill, C. D. & Hollenberg, L. C. Performance optimization for drift-robust fidelity improvement of two-qubit gates. *Phys. Rev. Appl.* **15**, 014023 (2021).

30. 30.

Blume-Kohout, R. et al. A taxonomy of small Markovian errors. Preprint at <https://arxiv.org/abs/2103.01928> (2021).

31. 31.

Cerfontaine, P., Otten, R., Wolfe, M. A., Bethke, P. & Bluhm, H. High-fidelity gate set for exchange-coupled singlet-triplet qubits. *Phys. Rev. B* **101**, 155311 (2020).

32. 32.

Pan, A. et al. Resonant exchange operation in triple-quantum-dot qubits for spin-photon transduction. *Quantum Sci. Technol.* **5**, 034005 (2020).

33. 33.

Zajac, D. M. et al. Resonantly driven CNOT gate for electron spins. *Science* **359**, 439–442 (2017).

34. 34.

Martinis, J. M. & Geller, M. R. Fast adiabatic qubit gates using only σ_z control. *Phys. Rev. A* **90**, 022307 (2014).

35. 35.

Nielsen, E. et al. Gate set tomography. *Quantum* **5**, 557 (2021).

36. 36.

Hempel, C. et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Phys. Rev. X* **8**, 031022 (2018).

37. 37.

Chow, J. M. et al. Detecting highly entangled states with a joint qubit readout. *Phys. Rev. A* **81**, 062325 (2010).

38. 38.

McClean, J. R. et al. OpenFermion: the electronic structure package for quantum computers. *Quantum Sci. Technol.* **5**, 034014 (2020).

39. 39.

Ganzhorn, M. et al. Gate-efficient simulation of molecular eigenstates on a quantum computer. *Phys. Rev. Appl.* **11**, 044092 (2019).

40. 40.

Mądzik, M. T. et al. Precision tomography of a three-qubit donor quantum processor in silicon. *Nature* **601**, 348–353 (2022).

41. 41.

Zheng, G. et al. Rapid gate-based spin read-out in silicon using an on-chip resonator. *Nat. Nanotechnol.* **14**, 742–746 (2019).

42. 42.

Schaal, S. et al. Fast gate-based readout of silicon quantum dots using Josephson parametric amplification. *Phys. Rev. Lett.* **124**, 067701 (2020).

43. 43.

Reed, M. *Entanglement and Quantum Error Correction with Superconducting Qubits*. PhD thesis, Yale Univ. (2013).

44. 44.

Russ, M. et al. High-fidelity quantum gates in Si/SiGe double quantum dots. *Phys. Rev. B* **97**, 085421 (2018).

45. 45.

Yang, Y.-C., Coppersmith, S. N. & Friesen, M. Achieving high-fidelity single-qubit gates in a strongly driven charge qubit with $1/f$ charge noise. *npj Quantum Inf.* **5**, 12 (2019).

46. 46.

Koski, J. V. et al. Strong photon coupling to the quadrupole moment of an electron in a solid-state qubit. *Nat. Phys.* **16**, 642–646 (2020).

47. 47.

Russ, M., Philips, S., Xue, X. & Vandersypen, L. M. K. The path to high fidelity multi-qubit gates for quantum dot spin qubits. *Bull. Am. Phys. Soc.* **66**, abstr. S29.00002 (2021).

48. 48.

Greenbaum, D. Introduction to quantum gate set tomography. Preprint at <https://arxiv.org/abs/1509.02921> (2015).

49. 49.

Nielsen, E., Blume-Kohout, R. J., Rudinger, K. M., Proctor, T. J., Saldyt, L. & USDOE, *Python GST Implementation (PyGSTi) v. 0.9*, Tech. Rep. PyGSTi (Sandia National Laboratories, 2019).

50. 50.

Nielsen, E. et al. Probing quantum processor performance with pyGSTi. *Quantum Sci. Technol.* **5**, 044002 (2020).

51. 51.

White, A. G. et al. Measuring two-qubit gates. *J. Opt. Soc. Am. B* **24**, 172–183 (2007).

52. 52.

Jamiolkowski, A. Linear transformations which preserve trace and positive semidefiniteness of operators. *Rep. Math. Phys.* **3**, 275–278 (1972).

53. 53.

Taube, A. G. & Bartlett, R. J. New perspectives on unitary coupled-cluster theory. *Int. J. Quantum Chem.* **106**, 3393–3401 (2006).

Acknowledgements

We acknowledge discussions with P. Cerfontaine, C. Bureau-Oxton, M. T. Madzik, A. Morello, J. Helsen, B. Terhal, M. Veldhorst and all the members of the spin qubit team, and technical assistance from O. Benningshof, M. Sarsby, R. Schouten and R. Vermeulen. This research was funded by the Dutch Ministry of Economic Affairs through the allowance for Top Consortia for Knowledge and Innovation (TKI) and the Army Research Office (ARO) under grant number W911NF-17-1-0274. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ARO or the US Government. The US Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

Author information

Affiliations

1. QuTech, Delft University of Technology, Delft, The Netherlands

Xiao Xue, Maximilian Russ, Nodar Samkharadze, Brennan Undseth, Amir Sammak, Giordano Scappucci & Lieven M. K. Vandersypen

2. Kavli Institute of Nanoscience, Delft University of Technology, Delft, The Netherlands

Xiao Xue, Maximilian Russ, Brennan Undseth, Giordano Scappucci & Lieven M. K. Vandersypen

3. Netherlands Organisation for Applied Scientific Research (TNO), Delft, The Netherlands

Nodar Samkharadze & Amir Sammak

Contributions

X.X. performed the experiment, with help from N.S. and B.U. M.R. developed the theory model and analysed the data with X.X. N.S. fabricated the quantum dot device. A.S. and G.S. designed and grew the Si/SiGe heterostructure. X.X. and L.M.K.V. conceived the project. L.M.K.V. supervised the project. X.X., M.R. and L.M.K.V. wrote the manuscript, with input from all authors.

Corresponding author

Correspondence to [Lieven M. K. Vandersypen](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

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 Two-qubit processes.](#)

Average gate infidelities, process matrices (PTMs) and error generators of the six quantum gates in the chosen gate set. These results are analysed by the pyGSTi package using maximum-likelihood estimation.

Extended Data Fig. 2 Single-qubit processes.

Average gate infidelities and process matrices (PTMs) of the identity gates (idle gates) and single-qubit X/Y gates in the subspace of the individual qubits. The individual PTMs are calculated from the PTMs in the two-qubit space (see [Methods](#)).

Extended Data Fig. 3 Bell states predicted from the quantum processes.

Top panels show the real part of the reconstructed density matrices of the four Bell states $\langle|\{\Psi\}^+\rangle| = (|01\rangle + |10\rangle)/\sqrt{2}$ (a), $\langle|\{\Psi\}^-\rangle| = (|01\rangle - |10\rangle)/\sqrt{2}$ (b), $\langle|\{\Phi\}^+\rangle| = (|00\rangle + |11\rangle)/\sqrt{2}$ (c) and $\langle|\{\Phi\}^-\rangle| = (|00\rangle - |11\rangle)/\sqrt{2}$ (d). The colour code is the same as in Fig. 4. Bottom panels show the quantum circuit used to reconstruct the Bell states. $\langle\{\{\text{rm}\{Z\}}\}_i\{\{\text{rm}\{Q\}}_i\}^2\rangle$ is a virtual π -rotation around the (\hat{z}) axis on the i th qubit, which is executed by a phase update on the microwave reference clock of the qubit and, therefore, is error-free. We numerically estimate the state fidelities to be 98.42% for the $|\Psi^+\rangle$ and $|\Psi^-\rangle$ states and 97.75% for the $|\Phi^+\rangle$ and $|\Phi^-\rangle$ states.

Extended Data Fig. 4 Initial gate calibrations.

a, Decomposition of single-qubit and two-qubit gates. After each microwave burst for single-qubit rotations, a corresponding phase correction is applied to each qubit. The CZ gate is implemented by a barrier voltage pulse applied to gate B (orange) and negative compensation pulses applied to gates LP (blue) and RP (red), with the same shape as the barrier pulse. Single-qubit phase corrections are then applied on each qubit to compensate the frequency detuning induced by electron movement in the magnetic field gradient. **b, c**, Calibration of phase corrections on Q_1 induced by a single-qubit gate applied on Q_2 (ϕ_{21} , b) and on Q_1 (ϕ_{11} , c). A relative phase shift, $2\phi_{21}$ ($2\phi_{11}$), is determined by interleaving the target gate (a $\pi/2$ rotation) and its inverse (a $-\pi/2$ rotation) on Q_2 (Q_1) in a Ramsey interference sequence. **d, e**, Calibration of phase corrections on each qubit after the CZ gate, using Q_1 (d) and Q_2 (e) as the control qubits, respectively. When the amplitude of the barrier pulse is perfectly calibrated, the two curves in each experiment should be out of phase by 180° . However, when the barrier pulse amplitude is calibrated such that one of the two experiments shows a 180° phase difference (d), the phase difference in the other calibration experiment always deviates by a few degrees. One possible explanation is that the optional π -rotation applied to the control qubit induces a small, off-resonance rotation on the other qubit, causing an additional phase on the target qubit to appear in the measurement due to the commutation relation of the Pauli operators.

Extended Data Fig. 5 Pulse optimization.

a, b, Full error generators for a CZ gate calibrated by conventional Ramsey sequences (**a**) and after improving the calibration using the information extracted from **a** (**b**), resulting in fidelities of 97.86% and 99.65%, respectively. **c, d**, Seven Hamiltonian errors (IX, IY, XI, YI, IZ, ZI and ZZ) extracted from the error generators shown in **a** (**c**) and **b** (**d**). Owing to the crosstalk-induced additional phases shown in Extended Data Fig. 4, errors IZ, ZI and ZZ occur systematically in conventional calibrations. Error bars indicate the 2σ confidence intervals computed using the Hessian of the loglikelihood function. **e, f**, Shapes of the barrier pulses (**e**) and their corresponding J envelopes (**f**) for a CZ gate before and after being corrected by GST. Since the Hamiltonian to generate a CZ gate is $H = (II + IZ + ZI - ZZ)/2$, the positive ZZ error shown in **c** is corrected by increasing the amplitude of the pulse. The IZ and ZI errors are corrected by decreasing the phase shifts θ_1 and θ_2 after the CZ gate. Hamiltonian errors in single-qubit gates are corrected similarly. The results presented in **b** and **d** are achieved in four loops of correction, with each loop correcting the parameters by approximately 70% of the measured deviation.

Extended Data Fig. 6 Workflow of the variational quantum eigensolver algorithm.

The qubit Hamiltonian is typically transformed from the molecular Hamiltonian by JW transformation or BK transformation by a classical processor (see [Methods](#)). A HF initial state is encoded into the qubit states according to JW or BK transformation and then transformed by the quantum processor into a parameterized ansatz state by considering single and double excitation in the molecule using the UCC theory. The expectation value of each individual Hamiltonian term is directly measured by partial state tomography. The expectation of the total energy is then calculated by the weighted sum of the individual expectations. The result is fed into a classical optimizer, which suggests a new parameterized ansatz state for the next run. This process is repeated until the expectation of the total energy converges.

Extended Data Table 1 Gate metrics

Supplementary information

[Peer Review File](#)

Rights and permissions

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

Reprints and Permissions

About this article

Cite this article

Xue, X., Russ, M., Samkharadze, N. *et al.* Quantum logic with spin qubits crossing the surface code threshold. *Nature* **601**, 343–347 (2022). <https://doi.org/10.1038/s41586-021-04273-w>

- Received: 01 July 2021
- Accepted: 22 November 2021
- Published: 19 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04273-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

Silicon qubits move a step closer to achieving error correction

- Ada Warren
- Sophia E. Economou

News & Views 19 Jan 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04273-w>

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

- Article
- [Published: 19 January 2022](#)

Precision tomography of a three-qubit donor quantum processor in silicon

- [Mateusz T. Mądzik](#) ORCID: orcid.org/0000-0003-4857-3257¹ na1 nAff9,
- [Serwan Asaad](#) ORCID: orcid.org/0000-0002-6407-4434¹ na1 nAff10,
- [Akram Youssry](#) ORCID: orcid.org/0000-0001-7301-5834^{2,3},
- [Benjamin Joecker](#) ORCID: orcid.org/0000-0003-0263-5440¹,
- [Kenneth M. Rudinger](#) ORCID: orcid.org/0000-0002-3038-4389⁴,
- [Erik Nielsen](#)⁴,
- [Kevin C. Young](#) ORCID: orcid.org/0000-0002-4679-4542⁵,
- [Timothy J. Proctor](#) ORCID: orcid.org/0000-0003-0219-8930⁵,
- [Andrew D. Baczewski](#) ORCID: orcid.org/0000-0001-8553-9934⁶,
- [Arne Laucht](#) ORCID: orcid.org/0000-0001-7892-7963^{1,2},
- [Vivien Schmitt](#)¹ nAff11,
- [Fay E. Hudson](#) ORCID: orcid.org/0000-0003-0134-3657¹,
- [Kohei M. Itoh](#) ORCID: orcid.org/0000-0002-8153-4893⁷,
- [Alexander M. Jakob](#)⁸,
- [Brett C. Johnson](#) ORCID: orcid.org/0000-0002-2174-4178⁸,
- [David N. Jamieson](#) ORCID: orcid.org/0000-0001-7733-6715⁸,
- [Andrew S. Dzurak](#) ORCID: orcid.org/0000-0003-1389-5096¹,
- [Christopher Ferrie](#)²,
- [Robin Blume-Kohout](#) ORCID: orcid.org/0000-0001-8134-948X⁴ &
- [Andrea Morello](#) ORCID: orcid.org/0000-0001-7445-699X¹

- 2976 Accesses
- 1 Citations
- 273 Altmetric
- [Metrics details](#)

Subjects

- [Quantum information](#)
- [Qubits](#)

Abstract

Nuclear spins were among the first physical platforms to be considered for quantum information processing^{1,2}, because of their exceptional quantum coherence³ and atomic-scale footprint. However, their full potential for quantum computing has not yet been realized, owing to the lack of methods with which to link nuclear qubits within a scalable device combined with multi-qubit operations with sufficient fidelity to sustain fault-tolerant quantum computation. Here we demonstrate universal quantum logic operations using a pair of ion-implanted ³¹P donor nuclei in a silicon nanoelectronic device. A nuclear two-qubit controlled-Z gate is obtained by imparting a geometric phase to a shared electron spin⁴, and used to prepare entangled Bell states with fidelities up to 94.2(2.7)%. The quantum operations are precisely characterized using gate set tomography (GST)⁵, yielding one-qubit average gate fidelities up to 99.95(2)%, two-qubit average gate fidelity of 99.37(11)% and two-qubit preparation/measurement fidelities of 98.95(4)%. These three metrics indicate that nuclear spins in silicon are approaching the performance demanded in fault-tolerant quantum processors⁶. We then demonstrate entanglement between the two nuclei and the shared electron by producing a Greenberger–Horne–Zeilinger three-qubit state with 92.5(1.0)% fidelity. Because electron spin qubits in semiconductors can be further coupled to other electrons^{7,8,9} or physically shuttled across different locations^{10,11}, these results establish a

viable route for scalable quantum information processing using donor nuclear and electron spins.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Operation of a one-electron–two-nuclei quantum processor.



Fig. 2: Tomography of nuclear Bell states.

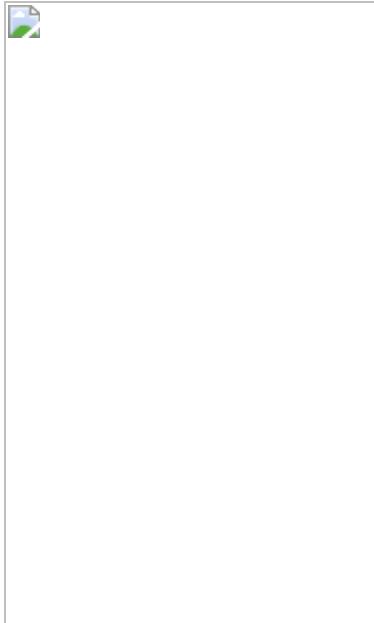


Fig. 3: Precise tomographic characterization of one- and two-qubit gate quality.

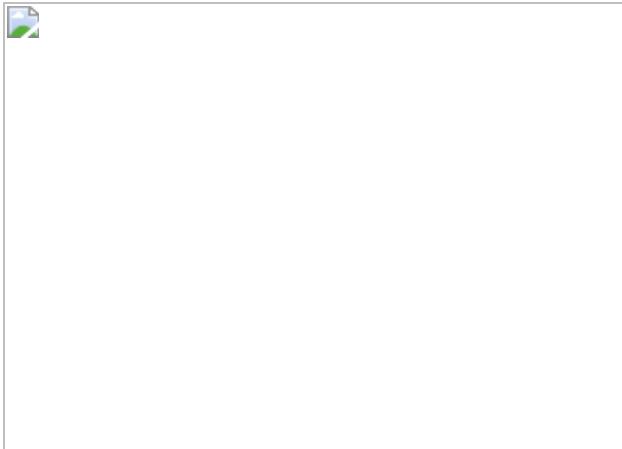
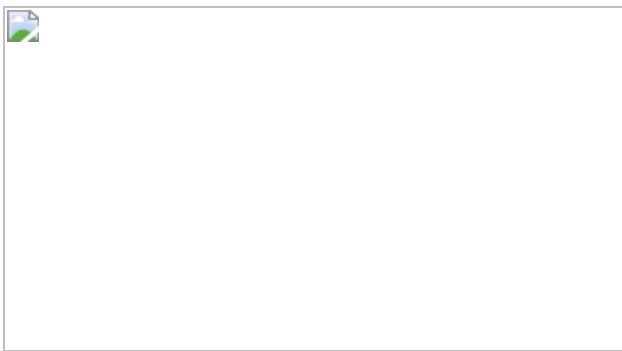


Fig. 4: Creation and tomography of an electron–nuclear three-qubit GHZ state.



Data availability

The experimental data that support the findings of this study are available in Figshare, <https://doi.org/10.6084/m9.figshare.c.5471706>. [Source data](#) are provided with this paper.

Code availability

The GST analysis was performed using a developmental version of pyGSTi that requires expert-level knowledge of the software to install and run. A future official release of pyGSTi will support the type of analysis performed here using a simple and well documented Python script. Until this code is available, interested readers can contact the corresponding author to get help with accessing and running the existing code. Multivalley effective mass theory calculations, some of the results of which are illustrated in Fig. 1b, were performed using a fork of the code first developed in the

production of ref. [60](#) that was extended to include multielectron interactions as reported in ref. [59](#). Requests for a license for and copy of this code will be directed to points of contact at Sandia National Laboratories and the University of New South Wales, through the corresponding author. The analysis code for Bell state tomography is in Figshare,
<https://doi.org/10.6084/m9.figshare.c.5471706>.

References

1. 1.
Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
2. 2.
Vandersypen, L. M. K. & Chuang, I. L. NMR techniques for quantum control and computation. *Rev. Mod. Phys.* **76**, 1037–1069 (2005).
3. 3.
Saeedi, K. et al. Room-temperature quantum bit storage exceeding 39 minutes using ionized donors in silicon-28. *Science* **342**, 830–833 (2013).
4. 4.
Filidou, V. et al. Ultrafast entangling gates between nuclear spins using photoexcited triplet states. *Nat. Phys.* **8**, 596–600 (2012).
5. 5.
Nielsen, E. et al. Gate set tomography. *Quantum* **5**, 557 (2021).
6. 6.
Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: Towards practical large-scale quantum computation.

Phys. Rev. A **86**, 032324 (2012).

7. 7.

Harvey-Collard, P. et al. Coherent coupling between a quantum dot and a donor in silicon. *Nat. Commun.* **8**, 1029 (2017).

8. 8.

He, Y. et al. A two-qubit gate between phosphorus donor electrons in silicon. *Nature* **571**, 371–375 (2019).

9. 9.

Madzik, M. T. et al. Conditional quantum operation of two exchange-coupled single-donor spin qubits in a MOS-compatible silicon device. *Nat. Commun.* **12**, 181 (2021).

10. 10.

Hensen, B. et al. A silicon quantum-dot-coupled nuclear spin qubit. *Nat. Nanotechnol.* **15**, 13–17 (2020).

11. 11.

Yoneda, J. et al. Coherent spin qubit transport in silicon. *Nat. Commun.* **12**, 4114 (2021).

12. 12.

Zhong, M. et al. Optically addressable nuclear spins in a solid with a six-hour coherence time. *Nature* **517**, 177–180 (2015).

13. 13.

Muhonen, J. T. et al. Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking. *J. Phys. Condens. Matter* **27**, 154205 (2015).

14. 14.

Bradley, C. et al. A ten-qubit solid-state spin register with quantum memory up to one minute. *Phys. Rev.* **9**, 031045 (2019).

15. 15.

Bourassa, A. et al. Entanglement and control of single nuclear spins in isotopically engineered silicon carbide. *Nat. Mater.* **19**, 1319–1325 (2020).

16. 16.

Waldherr, G. et al. Quantum error correction in a solid-state hybrid spin register. *Nature* **506**, 204–207 (2014).

17. 17.

Bhaskar, M. K. et al. Experimental demonstration of memory-enhanced quantum communication. *Nature* **580**, 60–64 (2020).

18. 18.

Pompili, M. et al. Realization of a multinode quantum network of remote solid-state qubits. *Science* **372**, 259–264 (2021).

19. 19.

Vandersypen, L. M. K. et al. Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent. *npj Quantum Inf.* **3**, 34 (2017).

20. 20.

Morello, A. et al. Single-shot readout of an electron spin in silicon. *Nature* **467**, 687–691 (2010).

21. 21.

Pla, J. J. et al. High-fidelity readout and control of a nuclear spin qubit in silicon. *Nature* **496**, 334–338 (2013).

22. 22.

Pla, J. J. et al. A single-atom electron spin qubit in silicon. *Nature* **489**, 541–545 (2012).

23. 23.

Ivie, J. A. et al. Impact of incorporation kinetics on device fabrication with atomic precision. *Phys. Rev. Appl.* **16**, 054037 (2021).

24. 24.

Hile, S. J. et al. Addressable electron spin resonance using donors and donor molecules in silicon. *Sci. Adv.* **4**, eaaq1459 (2018).

25. 25.

Anandan, J. The geometric phase. *Nature* **360**, 307–313 (1992).

26. 26.

James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001).

27. 27.

Dehollain, J. P. et al. Optimization of a solid-state electron spin qubit using gate set tomography. *New J. Phys.* **18**, 103018 (2016).

28. 28.

Blume-Kohout, R. et al. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nat. Commun.* **8**, 14485 (2017).

29. 29.

Huang, W. et al. Fidelity benchmarks for two-qubit gates in silicon. *Nature* **569**, 532–536 (2019).

30. 30.

Xue, X. et al. Benchmarking gate fidelities in a Si/SiGe two-qubit device. *Phys. Rev.* **9**, 021011 (2019).

31. 31.

Kimmel, S., da Silva, M. P., Ryan, C. A., Johnson, B. R. & Ohki, T. Robust extraction of tomographic information via randomized benchmarking. *Phys. Rev.* **4**, 011050 (2014).

32. 32.

Carignan-Dugas, A., Wallman, J. J. & Emerson, J. Bounding the average gate fidelity of composite channels using the unitarity. *New J. Phys.* **21**, 053016 (2019).

33. 33.

Blume-Kohout, R. et al. A taxonomy of small Markovian errors. Preprint at <https://arxiv.org/abs/2103.01928> (2021).

34. 34.

Proctor, T., Rudinger, K., Young, K., Sarovar, M. & Blume-Kohout, R. What randomized benchmarking actually measures. *Phys. Rev. Lett.* **119**, 130502 (2017).

35. 35.

Novais, E. & Mucciolo, E. R. Surface code threshold in the presence of correlated errors. *Phys. Rev. Lett.* **110**, 010502 (2013).

36. 36.

Neumann, P. et al. Multipartite entanglement among single spins in diamond. *Science* **320**, 1326–1329 (2008).

37. 37.

Takeda, K. et al. Quantum tomography of an entangled three-qubit state in silicon. *Nat. Nanotechnol.* **16**, 965–969 (2021).

38. 38.

Gullans, M. J. & Petta, J. R. Protocol for a resonantly driven three-qubit Toffoli gate with silicon spin qubits. *Phys. Rev. B* **100**, 085419 (2019).

39. 39.

Mehring, M., Mende, J. & Scherer, W. Entanglement between an electron and a nuclear spin 1/2. *Phys. Rev. Lett.* **90**, 153001 (2003).

40. 40.

Sackett, C. A. et al. Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).

41. 41.

Wei, K. X. et al. Verifying multipartite entangled Greenberger–Horne–Zeilinger states via multiple quantum coherences. *Phys. Rev. A* **101**, 032343 (2020).

42. 42.

Gross, J. A., Godfrin, C., Blais, A. & Dupont-Ferrier, E. Hardware-efficient error-correcting codes for large nuclear spins. Preprint at <https://arxiv.org/abs/2103.08548> (2021).

43. 43.

Asaad, S. et al. Coherent electrical control of a single high-spin nucleus in silicon. *Nature* **579**, 205–209 (2020).

44. 44.

Tosi, G. et al. Silicon quantum processor with robust long-distance qubit couplings. *Nat. Commun.* **8**, 450 (2017).

45. 45.

Pica, G., Lovett, B. W., Bhatt, R. N., Schenkel, T. & Lyon, S. A. Surface code architecture for donors and dots in silicon with imprecise and nonuniform qubit couplings. *Phys. Rev. B* **93**, 035306 (2016).

46. 46.

Buonacorsi, B. et al. Network architecture for a topological quantum computer in silicon. *Quantum Sci. Technol.* **4**, 025003 (2019).

47. 47.

Tosi, G., Mohiyaddin, F. A., Tenberg, S., Laucht, A. & Morello, A. Robust electric dipole transition at microwave frequencies for nuclear spin qubits in silicon. *Phys. Rev. B* **98**, 075313 (2018).

48. 48.

Mielke, J., Petta, J. R. & Burkard, G. Nuclear spin readout in a cavity-coupled hybrid quantum dot-donor system. *PRX Quantum* **2**, 020347 (2021).

49. 49.

Xue, X. et al. Quantum logic with spin qubits crossing the surface code threshold. *Nature* **601**, 343–347 (2022).

50. 50.

Noiri, A. et al. Fast universal quantum gate above the fault-tolerance threshold in silicon. *Nature* **601**, 338–342 (2022).

51. 51.

Adambukulam, C. et al. An ultra-stable 1.5 T permanent magnet assembly for qubit experiments at cryogenic temperatures. *Rev. Sci. Instrum.* **92**, 085106 (2021).

52. 52.

Kalra, R. et al. Vibration-induced electrical noise in a cryogen-free dilution refrigerator: characterization, mitigation, and impact on qubit coherence. *Rev. Sci. Instrum.* **87**, 073905 (2016).

53. 53.

Dehollain, J. et al. Nanoscale broadband transmission lines for spin qubit control. *Nanotechnology* **24**, 015202 (2012).

54. 54.

Feher, G. Electron spin resonance experiments on donors in silicon. I. Electronic structure of donors by the electron nuclear double resonance technique. *Phys. Rev.* **114**, 1219–1244 (1959).

55. 55.

Steger, M. et al. Optically-detected NMR of optically-hyperpolarized ^{31}P neutral donors in ^{28}Si . *J. Appl. Phys.* **109**, 102411 (2011).

56. 56.

Elzerman, J. M. et al. Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430**, 431–435 (2004).

57. 57.

Morello, A. et al. Architecture for high-sensitivity single-shot readout and control of the electron spin of individual donors in silicon. *Phys. Rev. B* **80**, 081307 (2009).

58. 58.

Braginsky, V. B. & Khalili, F. Ya. Quantum nondemolition measurements: the route from toys to tools. *Rev. Mod. Phys.* **68**, 1–11 (1996).

59. 59.

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).

60. 60.

Gamble, J. K. et al. Multivalley effective mass theory simulation of donors in silicon. *Phys. Rev. B* **91**, 235318 (2015).

61. 61.

Nielsen, E. et al. *Python GST Implementation (PyGSTi) v. 0.9*. Technical Report (Sandia National Lab, 2019).

62. 62.

Nielsen, E. et al. Probing quantum processor performance with pyGSTi. *Quantum Sci. Technol.* **5**, 044002 (2020).

63. 63.

Wilks, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **9**, 60–62 (1938).

64. 64.

Nielsen, E., Rudinger, K., Proctor, T., Young, K. & Blume-Kohout, R. Efficient flexible characterization of quantum processors with nested error models. *New J. Phys.* **23**, 093020 (2021).

65. 65.

Akaike, H. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike* (eds Parzen, E. et al.) 199–213 (Springer, 1998).

66. 66.

Tenberg, S. B. et al. Electron spin relaxation of single phosphorus donors in metal-oxide-semiconductor nanoscale devices. *Phys. Rev. B* **99**, 205306 (2019).

67. 67.

Hsueh, Y.-L. et al. Spin-lattice relaxation times of single donors and donor clusters in silicon. *Phys. Rev. Lett.* **113**, 246406 (2014).

Acknowledgements

We acknowledge conversations with W. Huang, R. Rahman, S. Seritan and C. H. Yang and technical support from T. Botzem. The research was supported by the Australian Research Council (grant no. CE170100012), the US Army Research Office (contract no. W911NF-17-1-0200), and the Australian Department of Industry, Innovation and Science (grant no. AUSMURI000002). We acknowledge support from the Australian National Fabrication Facility (ANFF). This material is based upon work supported in part by the iHPC facility at the University of Technology Sydney (UTS), by the US Department of Energy, Office of Science, Office of Advanced Scientific Computing Research’s Quantum Testbed Pathfinder and Early Career Research Programs, and by the US Department of Energy, Office of Science, National Quantum Information Science Research Centers (Quantum Systems Accelerator). Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and

Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the US Department of Energy, or the US Government.

Author information

Author notes

1. Mateusz T. Mądzik

Present address: QuTech, Delft University of Technology, Delft, The Netherlands

2. Serwan Asaad

Present address: Center for Quantum Devices, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

3. Vivien Schmitt

Present address: Université Grenoble Alpes, Grenoble INP, CEA, IRIG-PHELIQS, Grenoble, France

4. These authors contributed equally: Mateusz T. Mądzik, Serwan Asaad

Affiliations

1. School of Electrical Engineering and Telecommunications, UNSW Sydney, Sydney, New South Wales, Australia

Mateusz T. Mądzik, Serwan Asaad, Benjamin Joecker, Arne Laucht, Vivien Schmitt, Fay E. Hudson, Andrew S. Dzurak & Andrea Morello

2. Centre for Quantum Software and Information, University of Technology Sydney, Ultimo, New South Wales, Australia

Akram Youssry, Arne Laucht & Christopher Ferrie

3. Department of Electronics and Communication Engineering, Faculty of Engineering, Ain Shams University, Cairo, Egypt

Akram Youssry

4. Quantum Performance Laboratory, Sandia National Laboratories, Albuquerque, NM, USA

Kenneth M. Rudinger, Erik Nielsen & Robin Blume-Kohout

5. Quantum Performance Laboratory, Sandia National Laboratories, Livermore, CA, USA

Kevin C. Young & Timothy J. Proctor

6. Center for Computing Research, Sandia National Laboratories, Albuquerque, NM, USA

Andrew D. Baczewski

7. School of Fundamental Science and Technology, Keio University, Minato City, Yokohama, Japan

Kohei M. Itoh

8. School of Physics, University of Melbourne, Melbourne, Victoria, Australia

Alexander M. Jakob, Brett C. Johnson & David N. Jamieson

Contributions

M.T.M., V.S. and F.E.H. fabricated the device, with the supervision of A.M. and A.S.D., on an isotopically enriched ^{28}Si wafer supplied by K.M.I.

A.M.J., B.C.J. and D.N.J. designed and performed the ion implantation. M.T.M. and S.A. performed the experiments and analysed the data, with A.L. and A.M.’s supervision. B.J. and A.D.B. developed and applied computational tools to calculate the electron wavefunction and the Hamiltonian evolution. A.Y. designed the initial GST sequences, with C.F.’s supervision. K.M.R., E.N., K.C.Y., T.J.P. and R.B.-K. developed and applied the GST method. A.M., R.B.-K., M.T.M. and S.A. wrote the manuscript, with input from all coauthors.

Corresponding author

Correspondence to [Andrea Morello](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Christopher Wood 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 Device layout.](#)

Scanning electron micrograph of a device identical to the one used in this experiment. ^{31}P donor atoms are implanted in the region marked by the orange rectangle, using a fluence of $1.4 \times 10^{12} \text{ cm}^{-2}$ which results in a most probable inter-donor spacing of approximately 8 nm. Four metallic gates are fabricated around the implantation region, and used to modify the electrochemical potential of the donors. A nearby SET, formed using the SET top gate and barrier gates, enables charge sensing of a single donor atom, as well as its electron spin through spin-to-charge conversion (Methods). The tunnel coupling between the donors and SET is tuned by the rate gate situated between the SET and donor implant region. A nearby microwave (MW) antenna is used for ESR and NMR of the donor electron and nuclear spins, respectively.

Extended Data Fig. 2 Electrical tunability of the hyperfine interaction and the electron gyromagnetic ratio.

a, Map of the SET current as a function of SET gate and fast donor gates (pulsed jointly). The white dashed line indicates the location in gate space where the 2P donor cluster changes its charge state. The third, hyperfine-coupled electron is present on the cluster in the region to the right of the line. Electron spin readout is performed at the location indicated by the pink star. **b**, ESR spectrum of the electron bound to the 2P cluster, acquired while the system was tuned within the blue dashed rectangle in **a**. The hyperfine couplings A_1, A_2 are extracted from ESR frequencies as shown, namely $\langle\{A\}_1\rangle = (\{\nu\}_{\rm e}|\uparrow\downarrow\rangle + \{\nu\}_{\rm e}|\uparrow\uparrow\rangle)/2 - (\{\nu\}_{\rm e}|\downarrow\downarrow\rangle + \{\nu\}_{\rm e}|\downarrow\uparrow\rangle)/2$; $\langle\{A\}_2\rangle = \{\nu\}_{\rm e}(\uparrow\uparrow|\downarrow\downarrow\rangle - |\downarrow\downarrow|\uparrow\uparrow\rangle)$. **c, d**, Extracted hyperfine couplings within the marked area. The data show that A_1 decreases and A_2 increases upon moving the operation point towards higher gate voltages and away from the donor readout position. **e**, A small change is also observed in the sum of the two hyperfine interactions $A_t = A_1 + A_2$. **f**, Electrical modulation (Stark shift) of the electron gyromagnetic ratio γ_e , extracted from the shift of the average of the hyperfine-split electron resonances. The ESR frequencies can be tuned with fast donor gates at the rate of $\langle\Delta\nu\rangle$.

$$\begin{aligned} \langle \downarrow | \rm{e} \rangle \uparrow \uparrow = & 0.3 \langle \downarrow | \rm{MHzV} \rangle^{-1}; \langle \Delta \nu | \rm{e} \rangle \uparrow \uparrow \downarrow \downarrow = 5.2 \langle \downarrow | \rm{MHzV} \rangle^{-1}; \\ \langle \Delta \nu | \rm{e} \rangle \downarrow \downarrow \uparrow \uparrow = & 7.6 \langle \downarrow | \rm{MHzV} \rangle^{-1}; \langle \Delta \nu | \rm{e} \rangle \downarrow \downarrow \downarrow \downarrow = 2.4 \langle \downarrow | \rm{MHzV} \rangle^{-1}. \end{aligned}$$

Extended Data Fig. 3 Coherence metrics of the electron spin qubit.

The columns correspond to the nuclear configurations $\langle \downarrow \downarrow \rangle$, $\langle \downarrow \downarrow \uparrow \uparrow \rangle$, $\langle \uparrow \uparrow \downarrow \downarrow \rangle$, $\langle \uparrow \uparrow \uparrow \uparrow \rangle$, respectively. All measurements start with the electron spin initialized in the $\langle \downarrow \downarrow \rangle$ state. Error bars are 1σ confidence intervals. **a**, Electron Rabi oscillations. The measurements were performed by applying a resonant ESR pulse of increasing duration. The different Rabi frequencies f_{Rabi} on each resonance are probably due to a frequency-dependent response of the on-chip antenna and the cable connected to it. **b**, Electron spin-lattice relaxation times T_{1e} . Measurements were obtained by first adiabatically inverting the electron spin to $\langle \uparrow \uparrow \rangle$, followed by a varying wait time τ before electron readout. The observed relaxation times are nearly three orders of magnitude shorter than typically observed in single-electron, single-donor devices⁶⁶, and even shorter compared to $^{1\text{e}-2\text{P}}$ clusters. This strongly suggests that the measured electron is the third one, on top of two more tightly-bound electrons which form a singlet spin state⁶⁷. We also observe a strong dependence of T_{1e} on nuclear spin configuration. **c**, Electron dephasing times $\langle \{T\}_{2\{\rm{e}\}}^{\{\ast\}} \rangle$. The measurements were conducted by performing a Ramsey experiment—that is, by applying two $\pi/2$ pulses separated by a varying wait time τ , followed by electron readout. The Ramsey fringes are fitted to a function of the form $\langle \{P\}_{\uparrow}(\tau) \rangle = \{C\}_0 + \{C\}_1 \cos(\Delta\omega \tau + \Delta\phi) \exp[-\{(\tau / \{T\}_{2e})^{\{\ast\}}\}^2]$, where $\Delta\omega$ is the frequency detuning and $\Delta\phi$ is a phase offset. The observed $\langle \{T\}_{2\{\rm{e}\}}^{\{\ast\}} \rangle$ times are comparable to previous values for electrons coupled to a single ^{31}P nucleus. **d**, Electron Hahn-echo coherence times $\langle \{T\}_{2\{\rm{e}\}}^{\{\ast\}} \rangle$, obtained by adding a π refocusing

pulse to the Ramsey sequence. We also varied the phase of the final $\pi/2$ pulse at a rate of one period per $\tau = (5 \text{ kHz})^{-1}$, to introduce oscillations in the spin-up fraction which help improve the fitting. The curves are fitted to the same function used to fit the Ramsey fringes, with fixed $\Delta\omega = 5 \text{ kHz}$. The measured $(T_2 \{ \text{rm}\{e\} \})^{\text{last}}$ times are similar to previous observations for electrons coupled to a single ${}^3\text{P}$ nucleus.

Extended Data Fig. 4 Nuclear spin coherence times.

Panels in column 1 (2) correspond to nucleus Q1 (Q2). Error bars are 1σ confidence intervals. **a**, Nuclear dephasing times $(T_2 \{ \text{rm}\{n\} \})^{\text{last}}$, obtained from a Ramsey experiment. Results are fitted with a decaying sinusoid with fixed exponent factor 2 (see Extended Data Fig. 3). **b**, Nuclear Hahn-echo coherence times $(T_2 \{ \text{rm}\{n\} \})^{\text{Hahn}}$. To improve fitting, oscillations are induced by incrementing the phase of the final $\pi/2$ pulse with τ at a rate of one period per $(3.5 \text{ kHz})^{-1}$. Results are fitted with a decaying sinusoid with fixed exponent factor 2 (see Extended Data Fig. 3). **c**, Dependence of $(T_2 \{ \text{rm}\{n\} \})^{\text{Hahn}}$ on the amplitude of an off-resonance pulse. We perform this experiment to study whether a qubit, nominally left idle (or, in quantum information terms, subjected to an identity gate) is affected by the application of an RF pulse to the other qubit, at a vastly different frequency. Here, during the idle times between NMR pulses, an RF pulse is applied at a fixed frequency 20 MHz—far off resonance from both qubits’ transitions—with varying amplitude V_{RF} . The red dashed line indicates the applied RF amplitude for NMR pulses throughout the experiment. We observe a slow decrease of $(T_2 \{ \text{rm}\{n\} \})^{\text{Hahn}}$ with increasing V_{RF} . This is qualitatively consistent with the observation of large stochastic errors on the idle qubit, as extracted by the GST analysis in Fig. 3.

Extended Data Fig. 5 Nuclear spin quantum jumps caused by ionization shock.

The electron and nuclear spin readout relies upon spin-dependent charge tunnelling between the donors and the SET island. If the electron tunnels out of the two-donor system, the hyperfine interactions A_1, A_2 suddenly

drop to zero. If A_1 and A_2 include an anisotropic component (for example, due to the non-spherical shape of the electron wavefunction which results in nonzero dipolar fields at the nuclei), the ionization is accompanied by a sudden change in the nuclear spin quantization axes ('ionization shock'), and can result in a flip of the nuclear spin state. We measure the nuclear spin flips caused by ionization shock by forcibly loading and unloading an electron from the 2P cluster every 0.8 ms. **a**, For qubit 1 with $A_1 = 95$ MHz, the flip rate is $\langle\{\{\text{Gamma}\}\}_1\rangle = 2.8 \text{ times}$
 $\{10\}^{-6} \frac{\text{N}}{\text{rm}\{\text{flip}\}} \{N\}_{\text{rm}\{\text{ion}\}}$). **b**, For qubit 2 with $A_2 = 9$ MHz, the flip rate is $\langle\{\{\text{Gamma}\}\}_2\rangle = 4.0 \text{ times}$
 $\{10\}^{-7} \frac{\text{N}}{\text{rm}\{\text{flip}\}} \{N\}_{\text{rm}\{\text{ion}\}}$). This means that the nuclear spin readout via the electron ancilla is almost exactly quantum nondemolition. From this data, we also extract an average time between random nuclear spin flips of 283 seconds for qubit 1, and 2,000 seconds for qubit 2. The extremely low values of Γ —comparable to those observed in single-donor systems—are the reason why we can reliably operate the two ^{31}P nuclei as high-fidelity qubits.

Extended Data Fig. 6 CNOT and zero-CNOT nuclear two-qubit gates.

We perform Rabi oscillation on the control qubit followed by the application of **a**, zCNOT or **b**, CNOT gates. The two qubits are initialized in the $|(\Downarrow\Downarrow\rangle\langle\equiv|11\rangle)$ state. We observe the Rabi oscillations of both qubits in phase for zCNOT and out of phase for CNOT. At every odd multiple of $\pi/2$ rotation of the control qubit the Bell states are created.

Extended Data Fig. 7 Two-qubit GST.

a, Measurement circuit for the two-qubit GST. A modified version of this circuit has been used for Bell state tomography. The green box prepares the qubit 2 in the $|(\Updownarrow\rangle)$ state, then the orange box prepares the qubit 1 in the $|(\Updownarrow\rangle)$ state. The readout step in the blue box (see [Methods](#)) determines whether the $|(\Downarrow\Downarrow\rangle)$ state initialization was successful. Only then the record will be saved. The

electron spin is prepared in $\downarrow \rangle$ during the nuclear spin readout process. Subsequently, the GST sequence is executed. The red box indicates the Q1, Q2 readout step. The total duration of the pulse sequence is 120 ms, of which nuclear spin initialization is 8.6 ms (green and yellow), initial nuclear spin readout is 26.5 ms (blue), 3 ms delay is added for electron initialization (between blue and purple), GST circuit is 10 μ s–300 μ s (purple), and nuclear readout is 80 ms (orange). **b**, Measurement results for individual two-qubit GST circuit. The first 145 circuits estimate the preparation and measurement fiducials, and the subsequent circuits are ordered by increasing circuit depth. At the end of a circuit, there are three situations for the target state populations: 1) the population is entirely in one state, while all others are zero; 2) the population is equally spread over two states, while the other two are zero; 3) the population is equally spread over all four states. The measured state populations for the different circuits therefore congregate around the four bands 0, 0.25, 0.5, and 1, as indicated by black dashed lines.

[Extended Data Fig. 8 Estimated gate set, from process matrices to error rates.](#)

Experimental GST data were analysed using pyGSTi to obtain self-consistent maximum likelihood estimates of two-qubit process matrices for all six elementary gates. These are represented ('Process Matrix' column) in a gauge that minimizes their average total error, as superoperators in the two-qubit Pauli basis. Green columns indicate positive matrix elements, orange ones are negative. Wireframe sections indicate differences between estimated and ideal (target) process matrices. Those process matrices can be transformed to error generators ('Error Generator' column) that isolate those differences, and are zero if the estimated gate equals its target. Each gate's error generator was decomposed into a sparse sum of Hamiltonian and stochastic elementary error generators³³. Those rates are depicted ('All Error Rates' column) as contributions to the gate's total error, with 1σ uncertainties indicated in parentheses. Each non-vanishing elementary error rate (error generators are denoted 'H' or 'S' followed by a Pauli operator) is listed, and identified with its role in the total error budget (reproduced from Fig. 3). Orange bars indicate stochastic errors, dark blue indicate coherent errors that are intrinsic to the gate, and light blue indicate relational

coherent errors that were assigned to this gate. Total height of the blue region indicates the total coherent error, but because coherent error amplitudes add in quadrature, individual components' heights are proportional to their quadrature.

Extended Data Fig. 9 Simulation of standard and interleaved randomized benchmarking.

All simulated randomized benchmarking experiments used two-qubit Clifford subroutines compiled from the six native gates, requiring (on average) 14.58 individual gate operations per two-qubit Clifford. **a**, Standard randomized benchmarking, simulated using the GST-estimated gate set, yields a ‘reference’ decay rate of $r_r = 22.2(2)\%$, suggesting an average per-gate error rate of $r_r/14.58 \approx 1.5\%$. 1σ confidence intervals are indicated in parentheses. **b–f**, Simulated interleaved randomized benchmarking for the CZ gate, and one-qubit $X_{\pi/2}$ and $Y_{\pi/2}$ gates on each qubit, yielded interleaved decay rates $r_r + r_i$. For each experiment, 1,000 random Clifford sequences were generated, at each of 15 circuit depths m , and simulated using the GST process matrices. Exact probabilities (effectively infinitely many shots of each sequence) were recorded. Inset histograms show the distribution over 1000 random circuits at $m = 4$. Observed decays are consistent with each gate’s GST-estimated infidelities—for example, $1 - F = 0.79\%$ for the CZ gate (**b**). Performing these exact randomized benchmarking experiments in the lab would have required running 90,000 circuits to estimate a single parameter (r_i) for each gate to the given precision of $\pm 0.25\%$. Using fewer (<1,000) random circuits at each m would yield lower precision. GST required only 1,500 circuits to estimate all error rates to the same precision.

Extended Data Table 1 Estimated state preparation and measurement (SPAM) error rates

Supplementary information

Supplementary Information

This file contains supplemental information supporting the main claims of the paper. The information covers the following three components: supplemental data; extended GST analysis; and an analysis of possible causes for the observed two-qubit entangling errors.

[Peer Review File](#)

Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Mądzik, M.T., Asaad, S., Youssry, A. *et al.* Precision tomography of a three-qubit donor quantum processor in silicon. *Nature* **601**, 348–353 (2022). <https://doi.org/10.1038/s41586-021-04292-7>

- Received: 29 June 2021
- Accepted: 29 November 2021
- Published: 19 January 2022

- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04292-7>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [**Quantum logic with spin qubits crossing the surface code threshold**](#)
 - Xiao Xue
 - Maximilian Russ
 - Lieven M. K. Vandersypen

Nature (2022)

[**Silicon qubits move a step closer to achieving error correction**](#)

- Ada Warren
- Sophia E. Economou

News & Views 19 Jan 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04292-7>

- Article
- Open Access
- [Published: 19 January 2022](#)

Topological triple phase transition in non-Hermitian Floquet quasicrystals

- [Sebastian Weidemann](#) [ORCID: orcid.org/0000-0002-0090-1760¹](#),
- [Mark Kremer](#) [ORCID: orcid.org/0000-0003-2597-7259¹](#),
- [Stefano Longhi^{2,3}](#) &
- [Alexander Szameit](#) [ORCID: orcid.org/0000-0003-0071-6941¹](#)

Nature volume 601, pages 354–359 (2022)

- 2459 Accesses
- 37 Altmetric
- [Metrics details](#)

Subjects

- [Micro-optics](#)
- [Quantum simulation](#)

Abstract

Phase transitions connect different states of matter and are often concomitant with the spontaneous breaking of symmetries. An important category of phase transitions is mobility transitions, among which is the well known Anderson localization¹, where increasing the randomness induces a metal–insulator transition. The introduction of topology in condensed-matter physics^{2,3,4} lead to the discovery of topological phase transitions and materials as topological insulators⁵. Phase transitions in the symmetry of non-Hermitian systems describe the transition to on-average conserved energy⁶ and new topological phases^{7,8,9}. Bulk conductivity, topology and non-Hermitian symmetry breaking seemingly emerge from different physics and, thus, may appear as separable phenomena. However, in non-Hermitian quasicrystals, such transitions can be

mutually interlinked by forming a triple phase transition¹⁰. Here we report the experimental observation of a triple phase transition, where changing a single parameter simultaneously gives rise to a localization (metal–insulator), a topological and parity–time symmetry-breaking (energy) phase transition. The physics is manifested in a temporally driven (Floquet) dissipative quasicrystal. We implement our ideas via photonic quantum walks in coupled optical fibre loops¹¹. Our study highlights the intertwinement of topology, symmetry breaking and mobility phase transitions in non-Hermitian quasicrystalline synthetic matter. Our results may be applied in phase-change devices, in which the bulk and edge transport and the energy or particle exchange with the environment can be predicted and controlled.

[Download PDF](#)

Main

Phase transitions are defined as drastic changes of a system’s characteristics upon a small change of a single parameter. A classic example stems from chemistry, where changes between solid, liquid and gaseous phases can be induced by varying the temperature or pressure. In material sciences, the transition between the ferromagnetic and paramagnetic phases of magnetic materials at the Curie temperature is another fundamental example. The notion of phase transitions is established for understanding a diversity of different systems and phenomena, for instance, the evolution of planets in astrophysics¹², intracellular functioning¹³ and the emergence of disease in biological systems¹⁴, Bose–Einstein condensation¹⁵, and the evolution of the early Universe and the formation of the fundamental forces^{16,17,18}.

An important category among phase transitions—which are manifested in a plethora of different systems and phenomena—is mobility transitions^{19,20}. These are a drastic change of the conductance, that is, the spreading and localization of quantum particles, upon altering a parameter beyond a critical point. When certain metals and ceramics are cooled below a critical temperature, superconductivity abruptly emerges²¹. Another seminal example is Anderson localization¹, which is a sudden metal–insulator transition when uncorrelated randomness in a system is increased beyond a critical level.

A paradigmatic model showing a mobility transition is the Aubry–André–Harper (AAH) model²². It describes a one-dimensional system in an intermediate phase between perfect periodic order (crystal) and a completely disordered medium, possessing only long-range order without periodicity. This so-called quasicrystal²³ is known for undergoing a metal–insulator phase transition in one dimension at a critical value of the potential strength^{24,25}. At the critical point, the AAH model reduces to the

Harper equation²⁶ that can be directly mapped onto the two-dimensional Hofstadter model, which describes integer quantum Hall topology on a square lattice, resulting in the well known Hofstadter butterfly energy spectrum²⁷.

Metal–insulator phase transitions have usually been regarded as unrelated to other types of phase transition, such as spontaneous symmetry breaking occurring in dissipative systems or topological phase transitions observed in topological matter. This common wisdom has been challenged by recent theoretical studies^{9,10}, where the intriguing interplay between aperiodic order and dissipation has been unravelled. In such systems, the localization of the wavefunctions and the metal–insulator phase transition are usually associated with a spectral (symmetry breaking) phase transition²⁸, which can be characterized by the change of a topological (winding) number emerging from the closed contours of the eigenvalue spectrum in the complex plane⁹. In general, non-Hermitian contributions, such as dissipation, break the Hermitian time-reversal symmetry. However, one might find a combined parity–time (PT) symmetry⁶, which allows for a completely real energy spectrum, if unbroken. The discovered exotic behaviour at the critical point of PT phase transitions has sparked numerous applications, such as enhanced sensing²⁹, unidirectional invisibility³⁰ and mode-selective vortex lasing³¹.

Here we experimentally demonstrate a triple phase transition, where a topological phase transition, a mobility phase transition and spontaneous PT-symmetry breaking coincide (Fig. 1). We consider a non-Hermitian Floquet quasicrystal with PT symmetry, supporting the non-Hermitian skin effect^{11,32,33}, and possessing non-trivial point-gap topology⁹. Remarkably, the triple phase transition is observed by changing a single parameter, which can be purely Hermitian (strength of the nearest-neighbour coupling) or purely non-Hermitian (strength of the non-Hermitian gauge field), both of which we connect in a phase transition equation.

Fig. 1: Intertwinement of the triple phase transition in the non-Hermitian Floquet AAH model.

 **figure 1**

By intertwining a one-dimensional quasicrystal (AAH model) with a non-Hermitian anisotropy (skin effect model) via a temporal driving (Floquet mechanism), one obtains the non-Hermitian Floquet AAH model that is discussed in this work.

Although none of the models on the left show a phase transition, except for a localization transition in the AAH model, the combined model on the right shows a triple phase transition, that is, a localization phase transition and a topological phase transition are connected to spontaneous PT symmetry breaking (energy-conservation transition). All three transitions occur at the same critical point of a single parameter, which can be a purely Hermitian parameter (the coupling strength) or a purely non-Hermitian parameter (the anisotropy strength).

We implement our ideas in an optical system with controllable dissipation, consisting of coupled optical fibre loops^{11,34,35}, where the light propagation corresponds to the time evolution of a single-particle wavefunction within a one-dimensional discrete lattice. By modulating the phase of the light, we are able to emulate a quasicrystalline lattice potential. In the Hermitian aperiodic lattice, we observe a Floquet version of the Hofstadter butterfly quasienergy spectrum. When an imaginary gauge field is added to the quasicrystal, we observe a topological triple phase transition.

Theory

We start by introducing a Floquet version of the Hermitian AAH model and derive its metal–insulator transition, as well as its energy spectrum at the phase transition point (that is, the Floquet Hofstadter butterfly). In a second step, we include a non-Hermitian skin effect modulation¹¹ to observe the topological, PT symmetry-breaking and metal–insulator phase transitions.

In its original form, the AAH model²² represents a chain of sites with an on-site potential that is incommensurable compared with the lattice spacing, thus realizing a

quasicrystal²³. We consider a Floquet implementation of the AAH model, where the static AAH model is mapped onto a discrete-time quantum walk^{34,36}, as shown in Fig. 2. Such a quantum walk describes the evolution of a quantum particle on discrete lattice sites n , which are coupled to only the neighbouring sites, in a two-step protocol along the propagation direction m (time axis). The underlying lattice potential is controlled by introducing local phase terms between the couplings. The dynamics is governed by the equations

$$\begin{aligned} \text{\$}\$\{u\}_{n}^{m+1} &= (\{\cos(\beta)u\}_{n+1}^m + \{\mathrm{rm}\{i\}\}\sin(\beta) \\ &\quad \{v\}_{n+1}^m)\{\{\mathrm{rm}\{e\}\}\}^{\{\{\mathrm{rm}\{i\}\}\}}\{\varphi\}_{n}^u, \\ (1) \quad \text{\$}\$\{v\}_{n}^{m+1} &= (\{\mathrm{rm}\{i\}\}\sin(\beta)\{u\}_{n-1}^m + \{\cos(\beta)v\}_{n-1}^m) \\ &\quad \{\{\mathrm{rm}\{e\}\}\}^{\{\{\mathrm{rm}\{i\}\}\}}\{\varphi\}_{n}^v, \end{aligned}$$

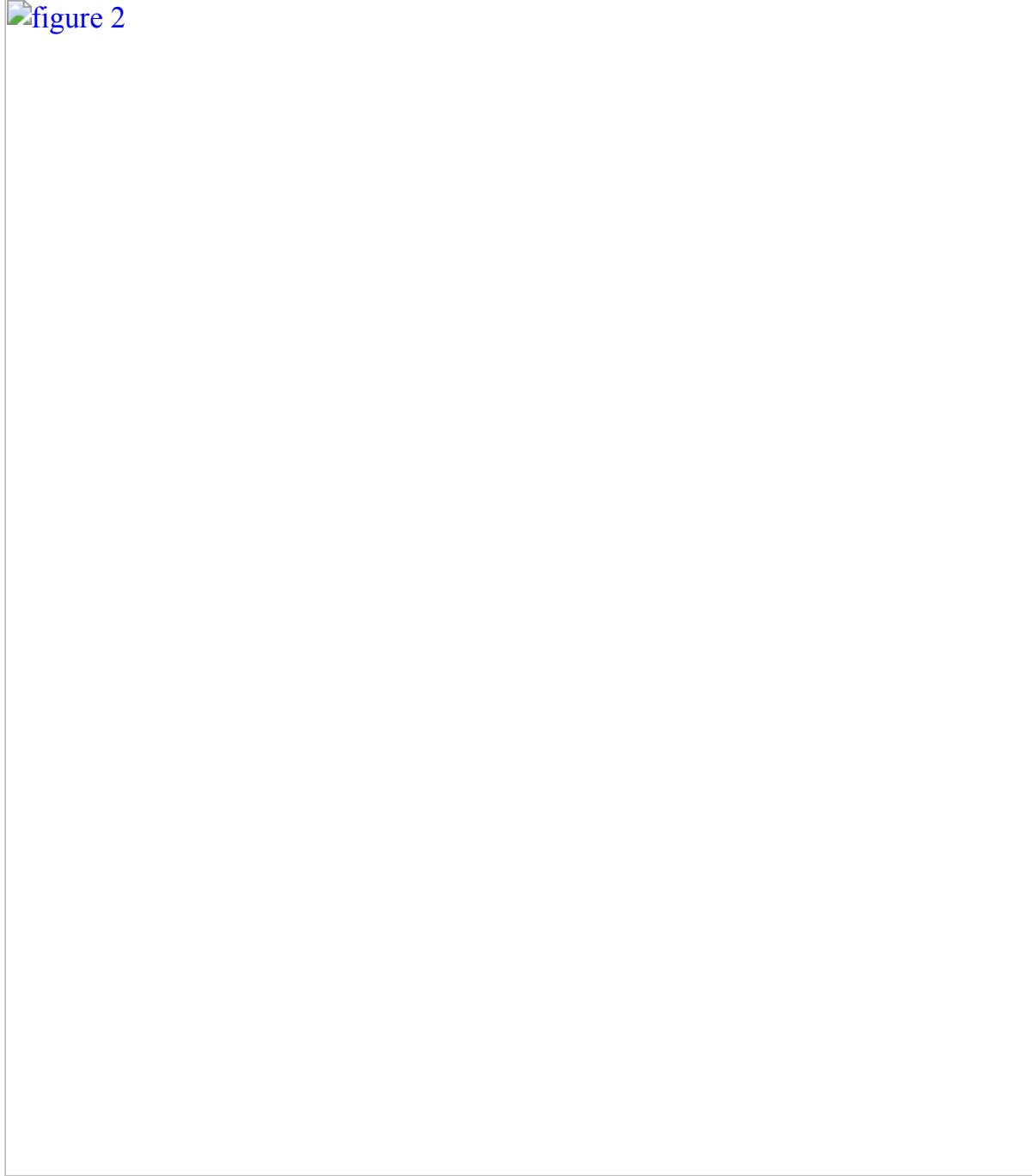
where $\langle\{u\}_n^m\rangle$ denotes the amplitude on left-moving paths of the quantum walk and $\langle\{v\}_n^m\rangle$ denotes the corresponding amplitude on right-moving paths at position n and time step m (Supplementary Section 1). We find that it is possible to implement a Floquet version of the AAH model by a temporally changing two-step phase modulation $\langle\{\varphi\}_n^{u,v}\rangle$ and constant sublattice (nearest-neighbour) coupling β . This modulation reads

$$\begin{aligned} \text{\$}\$\{\varphi\}_n^u &= \{\left(-1\right)^m\left(n+\frac{1}{2}\right)\frac{\{\mathrm{rm}\{\pi\}\}\phi}{2}, \\ (2) \quad \text{\$}\$\{\varphi\}_n^v &= \{\left(-1\right)^{m+1}\left(n-\frac{1}{2}\right)\frac{\{\mathrm{rm}\{\pi\}\}\phi}{2}, \end{aligned}$$

where ϕ is an irrational number for an incommensurable potential. The motivation for this modulation stems from the appearance of the vector potential in the original Harper model²⁶, where the magnetic field in the Landau gauge leads to phase gradients with opposite signs, depending on the hopping direction. This is reflected by the opposite signs in the phase terms for the left- and right-moving components $\langle\{\varphi\}_n^{u,v}\rangle$. The additional sign flip along the temporal direction m is based on the sublattice structure and assures that the same phase terms are aligned. This is illustrated in Fig. 2b, where the vertical lines connect identical phase modulations. Like in the original AAH model, our Floquet version displays a localization-delocalization phase transition without mobility edges, as the coupling parameter β is varied. The sudden metal-insulator phase transition is observed at the symmetry point $\beta = \pi/4$. This result can be rigorously proven from a self-similarity argument and it is also supported by a numerical analysis of the Lyapunov exponent (see Supplementary Section 2 for details). At this critical point, which lies exactly at the localization transition, when the phase gradient ϕ is varied, a Floquet version of the Harper–Hofstadter butterfly emerges for the quasienergy spectrum(Fig. 3b).

Fig. 2: Experimental implementation of one-dimensional quasicrystals via photonic quantum walks.

figure 2



a. Simplified experimental setting for realizing photonic lattices via discrete-time quantum walks. Two unequally long optical fibre loops are connected by a variable beam splitter (VBS) that controls β . The non-Hermitian potential is realized by introducing controlled optical losses with acousto-optical modulators (AOMs). A phase modulator (PM) emulates the real part of the potential and creates the

quasicrystalline order. Photodetectors measure the light intensity in both loops and hereby the time evolution of the quantum walk. **b**, The one-dimensional lattice (top) is implemented with a one-dimensional quantum walk (bottom) (equation (1)), based on a mesh lattice of beam splitters that is created with the coupled fibre loops (Supplementary Section 1). Gain and loss are incorporated at different lattice positions and in a two-step Floquet protocol, such that the skin effect modulation with anisotropic coupling with strength h (imaginary gauge field) is obtained. In a similar way, the potential (strength corresponds to amount of blue coloured filling) of the AAH model is realized via phase modulation (equation (2)) with a spatial phase gradient and alternating sign. The combined modulations realize the non-Hermitian Floquet AAH model based on a discrete-time quantum walk.

Fig. 3: Experimental Floquet metal–insulator transition and Floquet Hofstadter butterfly in the AAH model.

 figure 3

a, Upon a single-site excitation of the Hermitian quasicrystal, the spatial spreading of the wavefunction is shown via the second moment M_2 of its position operator after a large propagation time of $m = 200$. The grey area marks the tolerance region of expected deviations owing to limited accuracy in the lattice parameters (Methods).

One can see that a drastic spatial localization of the wavefunction sets in upon increasing the intersite coupling parameter beyond $\beta_c = \pi/4$ (top). This is exactly the metal–insulator phase transition, known from the static AAH model. **b**, At the symmetry point $\beta = \pi/4$, the Floquet Hofstadter–Harper model emerges. The evaluation of the quasienergies θ at $m = 380$ for 200 different phase gradients φ yields the Floquet Hofstadter butterfly (bottom) (Supplementary Section 3). The large propagation time allows for the high energy resolution. Compared with the original Hofstadter butterfly, our Floquet butterfly appears to be horizontally squeezed, due to the 2π periodicity in θ . The distribution of eigenvalues θ is obtained by applying the temporal Fourier transform (FT) to $\langle u_0^m \rangle$. Here, $\langle \left| u_0^m \right|^2 \rangle$ is retrieved from the intensity measurement. The phase information, which is only a \pm sign here, is lost in the intensity measurement, and we therefore added this minor information to the experimental data based on equation (1). The Floquet butterfly without this sign information is shown in Supplementary Section 3.

We now turn to the non-Hermitian extension of the AAH model. To this end, we employ a skin effect modulation¹¹. The exact mapping is shown in Fig. 2, where, besides a phase modulation, the amplitudes are also manipulated. Mathematically, this corresponds to a complex phase term $\langle \varphi_n^v \rangle \rightarrow \langle \varphi_n^v - \rm{i} h \rangle$ and $\langle \varphi_n^u \rangle \rightarrow \langle \varphi_n^u + \rm{i} h \rangle$, which, in turn, leads to an effective anisotropic coupling¹¹. The non-Hermitian contribution h to the phase term effectively corresponds to the implementation of an imaginary gauge field²⁸ (see Supplementary Section 5).

By superimposing the quasiperiodic phase potential with the non-Hermitian contribution, the non-Hermitian Floquet AAH model is formed. Our analysis shows that the non-Hermitian model exhibits a triple phase transition (Fig. 4a) at the critical point β_c , which is related to the anisotropy strength (or imaginary gauge field) h by (Supplementary Section 3)

$$h = \log \left(1 + \frac{1}{\cos(\beta_c)} \right) - \frac{1}{\cos \left(\frac{\pi}{4} \right)} \quad (3)$$

Fig. 4: Experimental triple phase transition in a non-Hermitian Floquet quasicrystal.

 figure 4

a, Three simultaneous phase transitions are shown from top to bottom. The two regimes of the triple phase transition are separated from left to right by the vertical dashed line. Top: for $\beta < \beta_c$ (left), the quasicrystal is expected to be in a topologically non-trivial phase, owing to the formation of point-energy gaps with non-zero winding $w = 1$. The quasienergy spectra are obtained from numerical diagonalization of the Floquet propagator with periodic boundary conditions. For $\beta > \beta_c$ (right), the topological phase changes, as the spectrum becomes real and the winding changes. Centre: for $\beta < \beta_c$ (left), the quasicrystal is in the broken PT phase, which is marked by the exponential growth λ of the overall energy $\langle \sum_n |u_n|^m \rangle^2 + \langle v_n^m \rangle^2 \propto e^{\lambda m}$ in time. For $\beta > \beta_c$ (right), the system changes to the unbroken PT phase, where the spectrum becomes real, and the overall energy becomes on-average constant. Bottom: for $\beta < \beta_c$ (left), the quasicrystal is in the delocalized phase, which is marked by a monotonic increase of the second moment that indicates strong spatial spreading of the wavefunction. For $\beta > \beta_c$ (right), all eigenstates become exponentially localized, which is marked by the extremely low and bounded second moment. All experiments are based on single-site excitations. The experimental data agree well with the predicted transition point at β_c

$= 0.275\pi$. The grey areas mark the tolerance regions of expected deviations owing to limited accuracy in the lattice parameters (Methods). **b**, Although a direct measurement of the winding number is not possible with the experimental setup, we observe light localization at a topological interface (top, $\beta_1 = 0.70\beta_c$ and $\beta_2 = 1.03\beta_c$ compared with a trivial interface (bottom, $\beta_1 = 0.70\beta_c$ and $\beta_2 = 0.89\beta_c$), where light does not localize at the interface. The localization at the interface vanishes, as soon as the right medium exceeds critical coupling β_c , such that both sides would have the same topological winding $w = 1$.

In the Hermitian limit $h \rightarrow 0$, the critical point takes the value of $\beta_c = \pi/4$, according to the self-duality argument. Besides the localization transition, now situated at $\beta_c(h)$, we find PT symmetry breaking at β_c , separating a regime with an exponential net energy growth ($\beta < \beta_c$) from a regime where the energy exchange with the environment is on-average balanced ($\beta > \beta_c$). A derivation of the PT symmetry is provided in Supplementary Section 6. The third transition that occurs at this point is a topological one. As the quasienergies of equation (1) form closed contours in the complex plane for $\beta < \beta_c$, a topological winding number can be introduced⁹

$$\text{w} = \frac{1}{2\pi} \int_{\partial D} \left(\frac{\partial \vartheta}{\partial \theta} \right) d\theta \quad (4)$$

which counts the number of times the complex spectral trajectory encircles a base point quasienergy θ_B when the phase θ is varied from zero to 2π . Here, H is the Hamiltonian with periodic boundary conditions and L is the overall number of sites (Supplementary Section 4). The phase θ adds to the phase modulation in the form $(\varphi_u \rightarrow \varphi_u + \vartheta / (2L))$ and $(\varphi_v \rightarrow \varphi_v - \vartheta / (2L))$ (Supplementary Section 4). One can further conclude from equation (3) that the triple phase transition shows a fundamental duality between Hermitian and non-Hermitian parameters, as it can be induced by changing either the site coupling β or the non-Hermitian gauge field h . Owing to the non-Hermitian contributions, the symmetry protection of the critical point is lost, and the phase transition point can be tuned with the anisotropy strength.

Experimental results

The Floquet non-Hermitian quasicrystal is implemented using an integrated photonic platform, in which classical light pulses propagate in two coupled and unequal long

fibre loops connected by a beam splitter (Fig. 2). This setting realizes a discrete-time quantum walk (Supplementary Section 1) and has proven to serve as versatile platform for the implementation of Hermitian^{37,38} and non-Hermitian^{11,35} synthetic lattices.

In a first experiment, the properties of the Hermitian Floquet AAH model are probed, using the potential of equation (2). Upon exciting a single lattice site $\langle\langle u\rangle_n^0, v\rangle_n^0\rangle = (\delta_{ij}, 0)$, where δ_{ij} denotes the Kronecker delta, we measure the evolving intensity distribution $\langle\langle u\rangle_n^m\rangle^2 + \langle\langle v\rangle_n^m\rangle^2$ of both loops. The localization transition is measured by evaluating the second moment $M_2 = \sum_n n^2 (\langle\langle u\rangle_n^m\rangle^2 + \langle\langle v\rangle_n^m\rangle^2)$ of the wave packet after $m = 200$ propagation steps, while using the irrational lattice frequency of $\phi = \left(\sqrt{5}-1\right)/2$, which is the inverse of the golden mean. The dynamical wave packet spreading is evaluated for several different splitting ratios β , encompassing the phase transition region around $\beta = \pi/4$. The results clearly show the occurrence of a delocalization–localization transition (Fig. 3a), in agreement with the numerical predictions.

To highlight the feasibility and controllability of our photonic setting, in a second experiment we measure the quasienergy spectrum at the phase transition point ($\beta = \pi/4$) for different potential frequencies in the range $(\phi \in [0, \pi])$, that is, different magnetic fluxes in the corresponding Harper equation. To obtain the information about the energy spectrum, we apply a Fourier transform on the lattice-site amplitudes along the propagation direction m (see Supplementary Section 3 for details). By using a large propagation time of $m = 380$, a high energy resolution is obtained. For the experiments, the phase modulation equation (2) is slightly adapted to the form $\langle\langle \varphi_n^u \rangle\rangle = (-1)^m n \phi$ and $\langle\langle \varphi_n^v \rangle\rangle = 0$, which is equivalent to the used modulation, as the lattice symmetry and the relative phase difference between the u and v components is maintained. With the adapted phase potential, a single site excitation $\langle\langle u\rangle_n^0, v\rangle_n^0\rangle = (\delta_{ij}, 0)$ now results in $\langle\langle u\rangle_0^m\rangle$ owing to the underlying symmetry of the lattice, which is explained in Supplementary Section 3. In an intensity measurement, any phase information is lost. However, as $\langle\langle u\rangle_0^m\rangle$, only the information about the sign is lost, which makes the phase measurement less crucial for the reconstruction of the spectrum. In Fig. 3b, we show the Fourier transform of the experimentally retrieved $\langle\langle u\rangle_0^m\rangle$, complemented with the sign information obtained from our simulations. The reconstructed energy spectra are reminiscent of the famous Hofstadter butterfly with a fractal structure, clearly showing that our Floquet photonic quantum walk can well reproduce the rich features of the Hermitian AAH model.

In the last and central experiment, we consider the non-Hermitian quantum walk, which displays the triple phase transition. The imaginary gauge phase h can be continuously varied from zero (Hermitian limit) to about $h \approx 0.12$. In the experiment, we fixed the gradient to the irrational Diophantine value $\phi = \left(\sqrt{5}-1\right)/2$. By varying the coupling β at a fixed non-vanishing gauge phase h , we simultaneously monitor the topology, energy exchange and bulk transport of the non-Hermitian quasicrystal (Fig. 4a). One can clearly see that the system undergoes three phase transitions as soon as β exceeds $\beta_c = 1.1\pi/4$, where all eigenvalues become real and the energy gaps in the complex plane are closed (Fig. 4a, top row). To verify the topological nature of this phase transition, we deduce the change of the winding number based on the dynamical (propagation) data, either in the bulk or by comparing a topologically non-trivial interface to a trivial interface (Fig. 4b). From the presence of a biased transport in the bulk owing to the non-Hermitian skin effect (as visible far from the interface), one can infer a non-zero winding³⁹, that is, a topologically non-trivial phase. However, the skin effect is suppressed in the localized phase, from which one can deduce a zero winding, that is, a topologically trivial phase^{9,39}. Furthermore, we numerically verified that only if the winding numbers of the adjacent media are different, light localizes at the interface instead of being transmitted through it. This observation supports the change of the topological phase, and such a behaviour suggests that the existing non-Hermitian bulk–boundary correspondence^{9,40} might be extendable to quasicrystals. Simultaneously, at the phase transition point, the exponential accumulation of energy in the systems stops, and the overall energy becomes on-average conserved as the PT phase transition from the broken to the unbroken PT phase occurs (Fig. 4a, centre row). In addition, a localization transition is observed: for $\beta < \beta_c$, there is a strong growth of the wavefunction’s second moment, whereas for $\beta > \beta_c$, the second moment becomes bounded, indicating that all eigenstates are localized (Fig. 4a, bottom row).

Our experimental results show that all three phase transitions coincide at $\beta_c(h)$. We stress that, although in the Hermitian limit only a localization–delocalization phase transition can be found at $\beta_c = \pi/4$, in the non-Hermitian case a triple phase transition emerges, where the critical point shifts to $\beta_c = 1.1\pi/4$, in agreement with our theoretical analysis.

Conclusion

We have experimentally demonstrated the concurrence of a triple phase transition in a one-dimensional non-Hermitian synthetic quasicrystal, which is realized in a Floquet photonic quantum walk with a controlled imaginary gauge field and an quasicrystalline potential. The usual metal–insulator phase transition found in the Hermitian limit corresponds to the simultaneous breaking of PT symmetry and to a

topological phase transition when a synthetic imaginary gauge field is applied to the quasicrystal. Our results provide experimental evidence on the exceptional properties of synthetic non-Hermitian quasicrystalline matter in terms of topology, localization-delocalization and symmetry-breaking phase transitions, which are responsible for phenomena that are drastically distinct from the familiar Hermitian realm. As such, our results have far-reaching consequences in a wide range of energy-conserving and open systems, as it offers a step towards unifying seemingly distinct phenomena.

Methods

Experimental setup

The experimental setup is similar to the setup described in refs. [11,35](#). The setup consists of two optical fibre loops, which are coupled by a variable beam splitter that controls the coupling parameter β . The loops are of unequal length, such that the roundtrip times are approximately given by $27 \mu\text{s} \pm 50 \text{ ns}$. The time difference of 100 ns defines the temporal width of a time-bin, in which the lattice positions n are encoded, such that approximately 270 positions can be encoded in the loops. The extended propagation time in each loop is achieved by using spools of single mode fibre (Corning Vascade LEAF EP). At the beginning of each measurement, a single 70-ns pulse is injected into the longer loop, here called the v-loop, via a fused fibre-optical beam splitter. The initial pulse is generated with a continuous-wave distributed feedback laser (JDS Uniphase, 1,550 nm) in combination with a Mach–Zehnder modulator (SDL Integrated Optics), which cuts out 70-ns pulses via intensity modulation. A pulse-picker acousto-optical modulator (Gooch & Housego) is used to further increase the on-off ratio of the light intensity. After the initial injection, the pulse circulates in the loop arrangement and periodically splits up at the variable beam splitter and multipath interference between the emerging subpulses takes place. The time multiplexing imposes the interference condition that two pulses will interfere only if they have travelled a permutation of the same sequence of long and short loop roundtrips. This interference condition guarantees an extremely stable phase relation between the interfering pulses, as external phase noise is acquired equally for a large frequency range of noise. The temporal intensity distribution of the propagating pulses is obtained by photodetectors (Thorlabs). The output voltages of the photodetectors are amplified with a logarithmic amplifier (FEMTO HLVA-100) and then acquired by an oscilloscope (R&S RTO1104). With the propagation timescales $\Delta t = 100 \text{ ns}$ and $T = 27 \mu\text{s}$, one can map the light intensity onto the discrete $1 + 1\text{D}$ lattice (time step m and position n). The measured pulse intensities correspond to the squared modulus of the wave function at lattice site n and time step m . To realize desired the phase and gain/loss modulation, an additional phase modulator (ixBlue Photonics) is placed in the u-loop and an acousto-optical amplitude modulator (zeroth order, Brimrose) is placed in each loop. To also realize gain and compensate for global losses (for

example, insertion loss or propagation losses in optical fibre) an erbium-doped fibre amplifier (Thorlabs) is placed in each loop. The amplifiers are optically gain clamped with by an additional distributed feedback laser (JDS Uniphase, 1,538 nm) that is coupled to the amplifier input via wavelength division multiplexing coupler (AC Photonics). Excess light from the gain clamping protocol is removed by an optical band-pass filter (WL Photonics), which also suppresses optical noise that stems from the amplification. All optical components are designed for operation at 1,550-nm wavelength and use a standard single mode fibre (SMF28 or comparable). The polarization is aligned between each loop and in front of polarization-sensitive components. Arbitrary waveform generators (Keysight Technologies, 33622A) generate the voltage signals that drive the electro-optical modulators. For each measurement, we perform an additional noise measurement in which no input pulse is injected. The measured light intensity then corresponds to the noise data and can be subtracted from the original date in post-processing.

Energy growth estimation

In the broken PT phase, it is expected that the overall light intensity can exponentially grow with propagation time m . Such growth in optical power can lead to a quick and nonlinear gain saturation of the amplifiers. Furthermore, high optical power might induce nonlinear effects via self-phase modulation or damage the optical components. To avoid the exponential light intensity growth, we impose artificial losses to the system in the broken phase, until the overall power no longer grows exponentially. These losses are equal in both loops and do not vary in time, such that the overall dynamic in the quantum walk is not affected. To do so, we decrease the gain of both amplifiers, such that an excess net loss is induced in each roundtrip. Afterwards, we keep these parameters and measure a Hermitian quantum walk as a control measurement. Owing to the excess loss, the overall light intensity of the Hermitian quantum walk is not constant, but exponentially decays with propagation time m . From this decay, one can deduce the power growth of the previously measured non-Hermitian system.

Experimental error and tolerance regions

The experimental error is captured via the systematic and statistical errors. The tolerance regions show the expected deviations owing to systematic errors. We assume that the main contribution for systematic errors stems from the limited precision of the experimentally realized lattice parameters. The lattice parameters are implemented via phase and amplitude modulation of the propagating light and via the coupling of the variable beam splitter that connects the fibre loops. Therefore, the systematic error stems mainly from the limited precision of the electro-optic driving, for instance, owing to bias drifts and tolerances in the look-up curves of the modulators. We assume

a relative error of $\pm 1\%$ in the modulation parameters (that is, the imprinted phases, gain/loss and the coupling β) and estimate the resulting error for the second moments and the energy growth via error propagation. The observed statistical fluctuations on repeating individual single-site excitations at least 10 times were negligible compared with the systematic error. The resulting overall errors can explain most of the discrepancies between the experimental and the theoretical data.

Accuracy of the irrational phase gradient

The experimental realization of the Hermitian and non-Hermitian Floquet AAH model required the phase gradient strength φ to be an irrational number to realize a potential that is incommensurable with respect to the lattice site spacing. It is therefore natural to ask with what accuracy an irrational parameter can be achieved and how any rational number (finite size) approximation would affect the results. On the one hand, the effect of a limited accuracy is already captured within the grey areas in Figs. 3, 4. Therefore, one can conclude that the limited accuracy of the phase modulation does not prevent an observation of the phase transitions that is based on the propagation data. On the other hand, it should be noted that owing to the time–energy uncertainty principle, in an experiment one can never resolve with an arbitrarily high precision the fine spectral or dynamical features observed when approaching the irrational φ . To clarify this point, let us assume that T is the largest observation time (time step) of the dynamics, and that we excite the lattice at the single site $n = 0$. The excitation spreads in the lattice with an upper speed bound $v \leq 1$ and thereby it is clear that one never probes more than $L = vT = T$ sites in the lattice at the left and the right sides from $n = 0$. Therefore, the experiment would yield the same results for two different values of the phase gradients φ_1 and φ_2 such that $d\varphi = |\varphi_2 - \varphi_1|$ is of the order of (or smaller than) $1/L = 1/T$, because the light pulses effectively probe the same potential over the spatial interval L . Hence, in the experiment, we cannot distinguish any finer spectral or dynamical features arising from any change of $d\varphi$ smaller than $1/T$, and this also sets the accuracy we require to achieve the target irrational value of $\varphi = (\sqrt{5}-1)/2$.

Data availability

All experimental data that have been used to produce the results reported in this manuscript are available in an open-access data repository⁴¹.

References

1. 1.

Anderson, P. W. Absence of diffusion in certain random lattices. *Phys. Rev.* **109**, 1492–1505 (1958).

2. 2.

Haldane, F. D. M. Model for a quantum Hall effect without landau levels: condensed-matter realization of the ‘parity anomaly’. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).

3. 3.

Thouless, D. J., Kohmoto, M., Nightingale, M. P. & Den Nijs, M. Quantized Hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982).

4. 4.

Kosterlitz, J. M. & Thouless, D. J. Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys. C* **6**, 1181–1203 (1973).

5. 5.

Hasan, M. Z. & Kane, C. L. Colloquium: Topological insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).

6. 6.

Bender, C. M. & Boettcher, S. Real spectra in non-hermitian hamiltonians having PT symmetry. *Phys. Rev. Lett.* **80**, 5243–5246 (1998).

7. 7.

Bergholtz, E. J., Budich, J. C. & Kunst, F. K. Exceptional topology of non-Hermitian systems. *Rev. Mod. Phys.* **93**, 015005 (2021).

8. 8.

Kawabata, K., Shiozaki, K., Ueda, M. & Sato, M. Symmetry and topology in non-Hermitian physics. *Phys. Rev. X* **9**, 041015 (2019).

9. 9.

Gong, Z. et al. Topological phases of non-Hermitian systems. *Phys. Rev. X* **8**, 031079 (2018).

10. 10.
Longhi, S. Topological phase transition in non-Hermitian quasicrystals. *Phys. Rev. Lett.* **122**, 237601 (2019).
11. 11.
Weidemann, S. et al. Topological funneling of light. *Science* **368**, 311–314 (2020).
12. 12.
Cheng, B., Mazzola, G., Pickard, C. J. & Ceriotti, M. Evidence for supercritical behaviour of high-pressure liquid hydrogen. *Nature* **585**, 217–220 (2020).
13. 13.
Choi, J.-M., Holehouse, A. S. & Pappu, R. V. Physical principles underlying the complex biology of intracellular phase transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
14. 14.
Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
15. 15.
Griffin, A., Snoke, D. & Stringari, S. *Bose–Einstein Condensation* (Cambridge Univ. Press, 1996).
16. 16.
Mazumdar, A. & White, G. Review of cosmic phase transitions: their significance and experimental signatures. *Rep. Prog. Phys.* **82**, 076901 (2019).
17. 17.
Linde, A. D. Phase transitions in gauge theories and cosmology. *Rep. Prog. Phys.* **42**, 389–437 (1979).
18. 18.

Kibble, T. W. B. Some implications of a cosmological phase transition. *Phys. Rep.* **67**, 183–199 (1980).

19. 19.

Sadovskii, M. V. Superconductivity and localization. *Phys. Rep.* **282**, 225–348 (1997).

20. 20.

Evers, F. & Mirlin, A. D. Anderson transitions. *Rev. Mod. Phys.* **80**, 1355–1417 (2008).

21. 21.

Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).

22. 22.

Serge Aubry, G. A. Analyticity breaking and Anderson localization in incommensurate lattices. *Ann. Isr. Phys. Soc.* **3**, 18 (1980).

23. 23.

Kraus, Y. E. & Zilberberg, O. Quasiperiodicity and topology transcend dimensions. *Nat. Phys.* **12**, 624–626 (2016).

24. 24.

Roati, G. et al. Anderson localization of a non-interacting Bose–Einstein condensate. *Nature* **453**, 895–898 (2008).

25. 25.

Lahini, Y. et al. Observation of a localization transition in quasiperiodic photonic lattices. *Phys. Rev. Lett.* **103**, 013901 (2009).

26. 26.

Harper, P. G. Single band motion of conduction electrons in a uniform magnetic field. *Proc. Phys. Soc. Sect. A* **68**, 874–878 (1955).

27. 27.

Hofstadter, D. R. Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields. *Phys. Rev. B* **14**, 2239–2249 (1976).

28. 28.

Hatano, N. & Nelson, D. R. Localization transitions in non-Hermitian quantum mechanics. *Phys. Rev. Lett.* **77**, 570–573 (1996).

29. 29.

Hodaei, H. et al. Enhanced sensitivity at higher-order exceptional points. *Nature* **548**, 187–191 (2017).

30. 30.

Lin, Z. et al. Unidirectional invisibility induced by PT-symmetric periodic structures. *Phys. Rev. Lett.* **106**, 213901 (2011).

31. 31.

Miao, P. et al. Orbital angular momentum microlaser. *Science* **353**, 464–467 (2016).

32. 32.

Helbig, T. et al. Generalized bulk–boundary correspondence in non-Hermitian topoelectrical circuits. *Nat. Phys.* **16**, 747–750 (2020).

33. 33.

Yao, S. & Wang, Z. Edge states and topological invariants of non-Hermitian systems. *Phys. Rev. Lett.* **121**, 086803 (2018).

34. 34.

Schreiber, A. et al. Photons walking the line: a quantum walk with adjustable coin operations. *Phys. Rev. Lett.* **104**, 050502 (2010).

35. 35.

Regensburger, A. et al. Parity–time synthetic photonic lattices. *Nature* **488**, 167–171 (2012).

36. 36.

Mochizuki, K., Kim, D. & Obuse, H. Explicit definition of PT symmetry for nonunitary quantum walks with gain and loss. *Phys. Rev. A* **93**, 062116 (2016).

37. 37.

Schreiber, A. et al. Decoherence and disorder in quantum walks: from ballistic spread to localization. *Phys. Rev. Lett.* **106**, 180403–180403 (2011).

38. 38.

Wimmer, M., Price, H. M., Carusotto, I. & Peschel, U. Experimental measurement of the Berry curvature from anomalous transport. *Nat. Phys.* **13**, 545–550 (2017).

39. 39.

Longhi, S. Probing non-Hermitian skin effect and non-Bloch phase transitions. *Phys. Rev. Res.* **1**, 023013 (2019).

40. 40.

Okuma, N., Kawabata, K., Shiozaki, K. & Sato, M. Topological origin of non-Hermitian skin effects. *Phys. Rev. Lett.* **124**, 086801 (2020).

41. 41.

Weidemann, S., Kremer, M., Longhi, S. & Szameit, A. Data set for topological triple phase transition in non-Hermitian Floquet quasicrystals. *Rostock University Publication Server* https://doi.org/10.18453/rosdok_id00003282 (2021).

Acknowledgements

We thank U. Peschel and M. Wimmer for discussions. S.W., M.K. and A.S. thank the Deutsche Forschungsgemeinschaft for funding their research (grants BL 574/13-1, SZ 276/9-2, SZ 276/15-1 and SZ 276/20-1) and the Krupp von Bohlen and Halbach foundation. S.L. acknowledges the Spanish State Research Agency through the Severo Ochoa and Maria de Maeztu Program for Centers and Units of Excellence in R&D (MDM-2017-0711)

Funding

Open access funding provided by Universität Rostock.

Author information

Author notes

1. These authors contributed equally: Sebastian Weidemann, Mark Kremer

Affiliations

1. Institute for Physics, University of Rostock, Rostock, Germany

Sebastian Weidemann, Mark Kremer & Alexander Szameit

2. Dipartimento di Fisica, Politecnico di Milano, Milan, Italy

Stefano Longhi

3. IFISC (UIB-CSIC), Instituto de Fisica Interdisciplinar y Sistemas Complejos - Palma de Mallorca, Palma, Spain

Stefano Longhi

Contributions

S.W. and M.K. designed the experimental implementation. S.W. performed the experiments. M.K. and S.L. developed the analytical model. A.S. supervised the project. All authors discussed the results and co-wrote the manuscript.

Corresponding author

Correspondence to [Alexander Szameit](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Yong Xu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

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

Supplementary information

Supplementary Information

This file contains Supplementary Notes 1–5, including Supplementary Figs. 1–4 and additional references.

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

Weidemann, S., Kremer, M., Longhi, S. *et al.* Topological triple phase transition in non-Hermitian Floquet quasicrystals. *Nature* **601**, 354–359 (2022).
<https://doi.org/10.1038/s41586-021-04253-0>

- Received: 22 June 2021
- Accepted: 09 November 2021
- Published: 19 January 2022

- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04253-0>

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/s41586-021-04253-0>

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

- Article
- Open Access
- [Published: 19 January 2022](#)

Chemical crystallography by serial femtosecond X-ray diffraction

- [Elyse A. Schriber](#) [ORCID: orcid.org/0000-0002-8935-2955](#)^{1,2} na1,
- [Daniel W. Paley](#) [ORCID: orcid.org/0000-0003-1161-5142](#)³ na1,
- [Robert Bolotovsky](#)³,
- [Daniel J. Rosenberg](#)^{3,4},
- [Raymond G. Sierra](#) [ORCID: orcid.org/0000-0002-6253-8282](#)⁵,
- [Andrew Aquila](#) [ORCID: orcid.org/0000-0003-0358-2774](#)⁵,
- [Derek Mendez](#)³,
- [Frédéric Poitevin](#) [ORCID: orcid.org/0000-0002-3181-8652](#)⁵,
- [Johannes P. Blaschke](#)⁶,
- [Asmit Bhowmick](#)³,
- [Ryan P. Kelly](#)^{1,2},
- [Mark Hunter](#) [ORCID: orcid.org/0000-0002-0110-7075](#)⁵,
- [Brandon Hayes](#)⁵,
- [Derek C. Popple](#)^{6,7},
- [Matthew Yeung](#)⁸,
- [Carina Pareja-Rivera](#) [ORCID: orcid.org/0000-0002-2764-0513](#)⁹,
- [Stella Lisova](#)¹⁰,
- [Kensuke Tono](#)¹¹,
- [Michihiro Sugahara](#)¹²,
- [Shigeki Owada](#)¹¹,
- [Tevye Kuykendall](#) [ORCID: orcid.org/0000-0003-1362-3285](#)¹³,
- [Kaiyuan Yao](#)¹⁴,
- [P. James Schuck](#) [ORCID: orcid.org/0000-0001-9244-2671](#)¹⁴,
- [Diego Solis-Ibarra](#) [ORCID: orcid.org/0000-0002-2486-0967](#)⁹,
- [Nicholas K. Sauter](#) [ORCID: orcid.org/0000-0003-2786-6552](#)³,
- [Aaron S. Brewster](#) [ORCID: orcid.org/0000-0002-0908-7822](#)³ &
- [J. Nathan Hohman](#) ^{1,2}

- 4684 Accesses
- 130 Altmetric
- [Metrics details](#)

Subjects

- [Characterization and analytical techniques](#)
- [Organic–inorganic nanostructures](#)

Abstract

Inorganic–organic hybrid materials represent a large share of newly reported structures, owing to their simple synthetic routes and customizable properties¹. This proliferation has led to a characterization bottleneck: many hybrid materials are obligate microcrystals with low symmetry and severe radiation sensitivity, interfering with the standard techniques of single-crystal X-ray diffraction^{2,3} and electron microdiffraction^{4,5,6,7,8,9,10,11}. Here we demonstrate small-molecule serial femtosecond X-ray crystallography (smSFX) for the determination of material crystal structures from microcrystals. We subjected microcrystalline suspensions to X-ray free-electron laser radiation^{12,13} and obtained thousands of randomly oriented diffraction patterns. We determined unit cells by aggregating spot-finding results into high-resolution powder diffractograms. After indexing the sparse serial patterns by a graph theory approach¹⁴, the resulting datasets can be solved and refined using standard tools for single-crystal diffraction data^{15,16,17}. We describe the ab initio structure solutions of mithrene (AgSePh)^{18,19,20}, thiorene (AgSPh) and tethrene (AgTePh), of which the latter two were previously unknown structures. In thiorene, we identify a geometric change in the silver–silver bonding network that is linked to its divergent optoelectronic properties²⁰. We demonstrate that smSFX can be applied as a general technique for structure determination of beam-sensitive microcrystalline materials at near-ambient temperature and pressure.

[Download PDF](#)

Main

Single-crystal X-ray diffraction (SCXRD) is a foundational characterization technique for chemistry and materials science, with over one million published organic and metal–organic structures^{2,3}. Frequently encountered challenges for SCXRD include difficulty in crystal growth, instability to atmosphere, solvent loss and radiation

sensitivity. Powder diffraction and electron microdiffraction are established methods for determining structures of microcrystalline substances, but each can be limited by some combination of the challenges above. As an example, the structure of the excitonic and blue-emitting mithrene, AgSeC_6H_5 , was determined by SCXRD^{18,19,20} but the remaining and optically divergent silver benzenechalcogenolates (AgEC_6H_5 , E = S, Te) have not yielded to characterization by Rietveld refinement or micro-electron diffraction²¹. In this work, we present the room-temperature (298 K) crystal structures of all three materials as determined by X-ray free-electron laser (XFEL) small-molecule serial femtosecond crystallography (smSFX). We first validate the method against the known structure of mithrene before presenting the previously unknown structures of the homologues thiorene, AgSC_6H_5 , and tethrene, AgTeC_6H_5 . Although the Ag–E coordination is similar in all three systems, we find a motif of linear Ag–Ag interactions in thiorene that provides insight into its lack of optical emission compared to the trigonal planar Ag–Ag network in the other two. smSFX is a general method with which to structurally characterize microcrystalline, low-symmetry and radiation-sensitive materials.

Electron diffraction and powder X-ray diffraction (PXRD) are currently the dominant techniques for structural studies on microcrystals. In PXRD work, peak broadening arises from sample-based and instrument-based factors and complicates the search for a unit cell and the measurement of structure factors. Therefore, Rietveld refinement is often more suitable for known structures or those with known homologues^{22,23}. When unknowns are characterized, the solution step is often model-assisted, for example, by simulated annealing^{24,25}. Three-dimensional electron diffraction (3D ED, also known as MicroED)^{4,5,6,7} and serial electron diffraction^{8,9,10} have also attracted attention for structure determination of nano- and microcrystals. Although this is generally performed under cryogenic conditions, some success has been reported using room-temperature 3D ED on radiation sensitive metal–organic frameworks with low-dose rotational scans¹¹. SerialED has been successful in characterizing structures of zeolites^{8,9} and proteins¹⁰. SerialED uses a single-shot-per-crystal approach, with sample preparation methods that mirror those of MicroED, where microcrystals are in vacuum conditions and affixed to a substrate. Here we present a complementary X-ray technique that also uses a single-shot-per-crystal approach. This method ensures intrinsic resistance to beam damage, requires almost no sample preparation, and is performed near room temperature and atmospheric pressure.

Serial femtosecond crystallography (SFX) usually involves a liquid jet of many (approximately 10^8) small crystals injected into the interaction point of an X-ray free-electron laser (XFEL). The high-flux, femtosecond pulses available at XFELs enable a ‘diffract before destruction’ approach, yielding Bragg reflections in single-shot frames from individual submicrometre crystals. The femtosecond duration of data collection

eliminates the role of beam damage²⁶. The data collection occurs at room temperature using a liquid jet of suspended crystals, preserving the integrity of samples that may be sensitive to freezing or desolvation. Sample preparation involves loading a sample suspended in solvent into a pump syringe that is attached to a jet delivery system. This can be sample consumptive, but there are adaptations available to conserve sample and even recirculate the suspension through the interaction point²⁷. High-speed data collection, currently in the kHz regime, is expected to give complete datasets in minutes of beam time²⁸.

There are several unique challenges for the application of SFX to chemical crystallography, generally centred on unit-cell determination and indexing diffraction patterns. A difference between traditional SCXRD and SFX is the unknown crystal orientation matrix. In a SCXRD experiment, the crystal is rotated on a goniometer, sequentially moving reflections into the diffracting condition. In SFX, each crystal is randomly oriented, so the orientation matrix must be determined independently for each frame (Fig. 1c). In macromolecular SFX, the large unit cell gives dozens of reflections per frame with obvious periodicity, permitting the use of Fourier methods for indexing^{29,30} (Extended Data Fig. 1). In contrast to macromolecular SFX, we find that typical frames for a hybrid material have only 3–10 reflections (Extended Data Fig. 2). This sparsity requires the use of indexing algorithms that do not rely on periodicity of the input reflections^{14,31}. We note that wider-band (pink-beam, Laue) methods can give more reflections per frame, a method applied to serial diffraction at synchrotron sources^{31,32}. In this work, we develop methods that use the narrow bandwidth of the intrinsic XFEL spectrum, about 0.3% $\Delta E/E$. Previous work¹⁴ has reported the partial structure of an amyloid peptide via SFX, applying the programme cctbx.small_cell to index small-molecule serial frames using a maximum clique algorithm that finds three-dimensional reciprocal space relationships in a sparse pattern. This algorithm requires a unit-cell candidate that must be discovered by other means.

Fig. 1: smSFX experiment and the analytes of interest.

 **figure 1**

a, Schematic of the smSFX experiment. **b**, Randomly oriented microcrystals are delivered to the XFEL interaction point. 30-fs XFEL pulses interact with the microcrystals to produce diffraction images before destruction by the XFEL pulse. **c**, Individual frames from randomly oriented mithrene crystals are sparse. **d–f**, Scanning electron micrographs of all three silver benzenechalcogenolates, where some size heterogeneity and morphological divergence is noted. Scale bars are 5 μm.

In this work we present a technique for synthesizing a high-resolution powder diffraction pattern from XFEL data before generating unit-cell candidates by a custom adaptation of the SVD-Index algorithm developed for powder diffraction^{33,34}. Using cctbx.small_cell indexing and room-temperature unit cells derived from the XFEL data, we index thousands of sparse partial diffraction patterns for mithrene, thiorene and tethrene. In all three cases we solve the crystal structures with no additional experimental information. The present work demonstrates that smSFX is a general technique for chemical crystallography on microcrystalline samples.

smSFX experiment

An overview of the experiment is presented in Fig. 1. Crystals are suspended in a solvent and subjected to the XFEL beam. We arrive at a complete diffraction dataset by measuring approximately 10^6 frames containing around 10^4 indexable patterns. The morphology of the microcrystalline powders of the three compounds is shown in the electron micrographs in Fig. 1d–f. Each compound was produced in stoichiometric yield and in a single procedural step by heating a mixture of silver (i) oxide and benzenethiol or benzeneselenol, or by reaction of silver powder with diphenyl ditelluride in isopropanol. Purified crystals were then held either dry or suspended in solvent until the experimental beamtime.

We performed three experiments at XFEL facilities for proof of concept, method validation and new science. The proof-of-concept smSFX experiment at the Linac Coherent Light Source (LCLS)¹² involved an attempt to refine the mithrene crystal structure against the Cuthbert model¹⁹ during a single 12-h shift. The resulting dataset permitted refinement with the known structure as a starting model, but an ab initio structure solution was not possible (further details are found in Methods). We then performed a follow-up experiment at SACLA¹³ where 4.5 h of data collection (after optimizing sample delivery) enabled an ab initio structure determination of mithrene from an smSFX dataset. A third experiment at LCLS was performed to obtain smSFX datasets from thiorene and tethrene microcrystals, materials with no known structure. The higher repetition rate at LCLS enabled complete datasets to be collected for thiorene and tethrene in approximately 3 h each. This experiment proved the versatility of smSFX across XFEL sources and its ability to determine structures of true unknowns.

Structure determination of mithrene

Unit-cell determination is the critical first step in solving a crystal structure, whether from SCXRD or SFX. In single-crystal diffraction, this task is accomplished using rotational scans and examining the periodicity of the lattice. In smSFX, the task is more difficult because no three-dimensional relationships between frames are immediately available. Following a previous approach¹⁴ we looked to the field of powder diffraction, where several algorithms are available for indexing a list of one-dimensional d spacings. As these techniques generally require highly accurate d values, we developed a method with which to derive high-resolution powder patterns by synthesizing data over numerous XFEL diffraction patterns (Fig. 2). We used dials.find_spots³⁵ to locate spots with subpixel accuracy (Fig. 2a), eliminating sample-based and instrument-based peak broadening. Full details are given in Methods; the synthetic powder diffraction patterns are presented in Fig. 2c–e. Candidate unit cells were generated by a custom multiprocessed interface to the SVD-Index algorithm as implemented in the popular powder diffraction suite GSAS-II³³. The candidate lattices

were evaluated by agreement with the powder pattern and by indexing rate for frames from the XFEL dataset. The correct unit cell showed good agreement with the synthetic powder pattern and had the highest indexing rate using the cctbx.small_cell algorithm (Fig. 3e and Extended Data Fig. 3). We therefore correctly determined the room-temperature monoclinic unit cell of mithrene in agreement with the literature single-crystal structure¹⁹.

Fig. 2: Powder patterns derived from XFEL data.

 **figure 2**

a, b, A histogram over thousands of individual Bragg spot observations identifies the d spacings of the lattice, expressed here as the diffraction angle 2θ (in $^{\circ}$). As described in Methods section ‘Synthetic powder patterns from XFEL shots’, first the spots are found from the full dataset, then the spot positions are converted to d spacings. The d

spacings are then binned into a histogram that amounts to a sharpened powder diffraction pattern. **c–e**, Radial averages of thiorene (**c**), mithrene (**d**) and tethrene (**e**) using the spot-finder radial histogram (black) compared to laboratory source XRD patterns (coloured). The horizontal axis is scaled to the wavelength of Cu K α radiation ($\lambda = 1.5406 \text{ \AA}$), according to Bragg's law ($\lambda = 2d\sin\theta$), and the vertical axes are shown in arbitrary units. Thiorene was affected by an unknown impurity in low concentration and thus weak reflections have been downweighted. Insets show magnification of the data around $2\theta = 37^\circ\text{--}43^\circ$.

Fig. 3: Mithrene structure determination results.

 figure 3

a, b, Identification of the C₆H₅ group in a Fourier difference map. The map is calculated from $F_{\text{obs}} - F_{\text{calc}}$ coefficients and contours are drawn at intervals of approximately 0.15 e⁻ per Å³. The ring is disordered over two positions; both positions

are shown in **b**, although they nearly overlap in this view. The plots were generated by Olex2¹⁵. **c, d**, Crystal structure of mithrene as determined in the present work (**c**) and the reference single-crystal structure¹⁹ (**d**). **e**, Plot of candidate unit cells indexing rate versus M^* (see equation (1) in Methods). Cell 2, labelled in red, is the correct cell, with the highest indexing rate and a low value of M^* .

Having determined the unit cell, we performed indexing, integration and merging using cctbx.small_cell_process and cctbx.xfel.merge. The space group was assigned by systematic absences and database frequency as is typical in chemical crystallography. Structure solution was performed by charge-flipping as implemented in olex2.solve¹⁵. This yielded a successful solution that revealed the heavy-atom substructure with Ag and Se distinguishable by their peak heights. At this stage, we discovered a simulated ‘twin operator’ that arose from a coincidental, pseudo-orthorhombic metric of the lattice. To resolve this, a second round of merging was performed using the heavy-atom partial structure as a reference for reindexing and scaling.

The model was completed by standard methods, using successive rounds of difference Fourier map calculation and atom assignment in ShelXL¹⁶. We were able to identify carbon atoms in difference maps (Fig. 3a, b). The C₆H₅ group was located in two positions disordered by rotation around the Se–C bond axis. Upon completion of the model, we performed one final round of merging with this complete SFX-determined structure as a reference for scaling and reindexing. The final refinement converged with $R_1(I > 2\sigma(I)) = 11.4\%$, where I -values are structure factor intensities and $\sigma(I)$ is the estimated uncertainty on I . Complete experimental and refinement statistics are given in Table 1.

Table 1 Crystallographic results for thiorene, mithrene and tethrene

Our approach of iterative structure refinement and re-scaling calls for some caution because there is a risk of biasing the data to match a model determined from a previous round of refinement. We demonstrated empirically (details in Methods) that the risk of model bias is not realized in these examples. In short, when a reference structure is ‘sabotaged’ by moving one atom a substantial distance, the final refinement returns that atom to its correct location; demonstrating that the intensities are not severely biased by the details of the reference model.

We validated the smSFX results by comparison to the literature structure determined by SCXRD (Fig. 3c, d and Table 1). The overall structures, including the disordered phenyl ring, are identical. The Ag–Se bonding distances were used as a quality metric. The mean deviation from the reference was 0.015 Å, which is comparable to the least-squares uncertainty on the distances in the smSFX structure (Extended Data Table 1).

Therefore, we conclude that the experimental errors are small and accurately estimated.

Crystal structure of thiorene

Following the approach established for mithrene, we identified a base-centred monoclinic lattice for thiorene (Table 1). The lattice was similar to that of mithrene, but with the a and b axes exchanged. This suggested that the structure would be similar, with the symmetry elements differently placed. After indexing and merging, we solved the structure by charge-flipping in olex2.solve, initially in centrosymmetric $C2/c$. The basic structural motif, as predicted from the unit cell, is a distortion of the layered mithrene structure with SC_6H_5 groups each coordinating four Ag atoms. After extensive model-building and revision, with details given in Methods, we identified an ordered, pseudo-centrosymmetric model in polar Cc with independent geometries for the top and bottom sides of each Ag–S sheet. The final refinement converged with R_1 ($I > 2\sigma(I)$) = 10.5% for data within the 1.35-Å limit imposed by the experimental geometry (Fig. 4).

Fig. 4: Crystal structures, Ag–Ag motifs and optical properties of all three benzenechalcogenolates.

 **figure 4**

a–c, Suspended microcrystals of thiorene (**a**), mithrene (**b**) and tethrene (**c**) show their respective milky white, yellow and deep orange colours. **d–f**, Side and top views of crystal structures from smSFX for thiorene (**d**), mithrene (**e**) and tethrene (**f**). Thermal ellipsoids for Ag (blue), S (yellow), Se (orange) and Te (magenta) are drawn at the

50% probability level. Hydrogen atoms and one position of disordered C₆H₅ (for mithrene) are omitted for clarity. **g–i**, Models of thiorene (**g**), mithrene (**h**) and tethrene (**i**) with the view oriented down the *c* axis of the unit cell, with the carbon and hydrogen atoms omitted, displaying the divergence in the thiorene Ag–Ag bonding environment compared to that in mithrene and tethrene. **j**, Photoluminescence spectra of mithrene and tethrene under 467-nm excitation. Mithrene displays a sharp photoluminescence peak at approximately 2.7 eV; tethrene displays a broadband photoluminescence peak in the approximate range 1.5–2.4 eV. No photoluminescence is observed for thiorene. **k**, Time-dependent photoluminescence for mithrene and tethrene, with lifetimes of 50 ps and 1 ns respectively. IRF, instrument response function.

Crystal structure of tethrene

We found a base-centred monoclinic lattice (Table 1), closely resembling the lattice of mithrene and thus probably isomorphous. After indexing and merging, we searched for solutions in *C*2/*c* using ShelXT¹⁷ and obtained a solution for the heavy-atom substructure. As expected, the structure closely resembles mithrene, with small distortions to accommodate the longer Ag–Te bonds. The model was completed using a combination of difference maps and rigid fragments; full details are given in Methods. The final refinement had $R_1(I > 2\sigma(I)) = 10.6\%$ for data within 1.35-Å resolution (Fig. 4).

Divergent optical behaviour explained

Mithrene and tethrene both have strong light–matter interactions; yellow mithrene has a direct bandgap and emits blue light with a short photoluminescent lifetime (approximately 50 ps);¹⁸ whereas orange tethrene has an indirect bandgap³⁶ and emits broadly with a longer lifetime (approximately 1 ns) (Fig. 4j, k). By contrast, white thiorene has no emissive profile in the visible range. Our results give a concrete structural explanation for these observations. We have previously demonstrated that excitons in mithrene, delocalized in two dimensions across the argentophilic network of Ag–Ag bonds, give rise to its visible absorption and emission spectra^{18,21,37}. The argentophilic interactions of thiorene are marked by linear Ag–Ag chains, which do not support the two-dimensional delocalization of excitons observed in mithrene and tethrene. Therefore, this smSFX study resolves the long-standing puzzle of the optoelectronic divergence of thiorene from its homologues mithrene and tethrene.

Conclusion

XFEL smSFX can be used to solve structures of obligate microcrystals. We demonstrated this approach by using the cctbx.small_cell programme for indexing and integration of single-shot images from data collected on three microcrystalline materials. The resulting datasets were used to solve a 1.2-Å structure of mithrene, using a previous structure¹⁹ only as a reference for method validation, and the 1.35-Å structures of thiorene and tethrene, with no reference structures. This approach should be broadly applicable across XFELs and synchrotron radiation facilities equipped for serial crystallography. Access to higher beam energies, with 25 keV expected at LCLS in 2022, will make resolutions of less than 0.7 Å available. Our results establish smSFX as a technique for microcrystal chemical crystallography that is free from beam damage. The intrinsically time-resolved nature of XFEL diffraction also presents opportunities for expansion into advanced structural studies.

Note added in Proof: After a recent report³⁸ of mithrene in $P2_1/c$, we checked our three datasets for possible primitive lattices. We confirmed that mithrene and tethrene are *C*-centred at room temperature. Cuthbert also reported a *C*-centred lattice for mithrene at 180 K (ref. ¹⁹). Taken together, these observations imply a possible symmetry-lowering phase transition below 180 K in mithrene. For thiorene, there are weak violations of the *C*-centring condition, but we could not refine a primitive model using the systematically weak $h + k = 2n + 1$ reflections. Data collected at atomic resolution in the upcoming LCLS runs may yield further insights.

Methods

Synthesis

Benzeneselenol (97%, 375152), diphenyl ditelluride (98%, 384127), silver (I) oxide (99%, 221163), silver powder (>99.9%, 327077), and thiophenol (97%, T32808) were used as received from Sigma Aldrich. Microcrystalline silver benzeneselenolate (mithrene) and its homologues silver benzene thiolate (thiorene) and silver benzenetellurolate (tethrene) were prepared in an approach adapted from our earlier report³⁹. Safety precaution: benzenethiol and benzeneselenol are toxic and foul-smelling, so adequate engineering protocols should be in place to prevent exposure. Benzeneselenol is susceptible to oxidation in air. Pure material should appear faintly yellow, whereas the presence of oxidized products (for example, diphenyl diselenide) will give a deeper orange colour, which generally will not impact the reaction until a substantial quantity of the benzeneselenol has oxidized. We endeavoured to plan our reactions such that the entire quantity of the benzeneselenol (1 g) was consumed, to eliminate the necessity of storing the opened containers.

For small-scale reactions producing mithrene and thiorene, 1 equivalent of silver (i) oxide (~50–150 mg) was placed in a 1-dram glass vial with 1.5 equivalents of benzeneselenol or benzenethiol. Reactions can be run solvent-free or in 1 ml of isopropanol. We find the addition of isopropanol is useful for subsequent sample recovery. The vial was sealed and heated in an oven to 70 °C. At elevated temperatures, reactions can appear complete within minutes; however, samples are allowed to incubate for 1 day to ensure stoichiometric yield. Caution when opening the vial is advised as the reaction is performed with excess benzeneselenol or benzenethiol, both of which are highly odorous and may be under pressure when hot. For our SACLA experiment, we used 3.0 g of silver oxide again with 1.5 equivalents of benzeneselenol in a 20 ml scintillation vial. The dramatic volume expansion of the silver oxide to the silver benzeneselenolate product caused the material to become impacted in the vial, so larger glassware may be advisable. Excess organic ligand is removed by 3 cycles of centrifugation (~10,000 rpm for 10 min), dilution with fresh isopropanol, and decanting. To prepare tethrene, 3:1 molar ratio of diphenyl ditelluride to solid silver (0) was added to a 4-dram vial with a PTFE-lined cap, along with 1 ml of isopropanol. The contents were thoroughly mixed and incubated at 80 °C in an oven for 4 days and recovered like the other examples. Product is transferred to a 45 ml Falcon tube and diluted with isopropanol. Products should not have a discernible odour. Crystals are stored refrigerated, when possible, protected from light, and either dry or suspended in methanol, ethanol or isopropanol.

Product purity was assessed by taking aliquots of the crystals from a slurry and depositing them on silicon substrates until dry. Laboratory powder X-ray diffraction experiments were sufficient to match powder patterns of earlier reports and to determine whether there was any silver oxide remaining. Images of all three benzenchalcogenolate crystals were imaged using a field-emission scanning electron microscope at an accelerating voltage of 1–5 kV.

smSFX at SACLA

Having prepared of the order of 10^{18} crystals, we used a Rayleigh jet injector system⁴⁰ at a high flow rate of $\sim 150 \mu\text{l min}^{-1}$ to reduce the possibility of clogging the capillary lines. The jet was positioned to intersect the 1 μm XFEL beam in atmospheric conditions at room temperature (25 °C). Diffraction patterns were recorded on the MPCCD detector⁴¹ using 12 keV (1.032 Å) X-ray pulses with a duration of ~ 30 fs at a frequency of 30 Hz. We increased concentration in a stepwise fashion from 0.2 mg ml^{-1} to 2 mg ml^{-1} until we observed a hit rate of $\sim 10\%$, which satisfied data-processing redundancy needs without causing capillary line clogs.

smSFX at LCLS

Thiorene and tethrene were transported as dry solids to the Coherent X-ray Imaging endstation at the Linac Coherent Light Source¹² for smSFX experiments. Solvated samples were delivered to the XFEL interaction point via a 3D-printed polymer Gas Dynamic Virtual Nozzle (GDVN)⁴² in high vacuum (10^{-8} torr). During prior experiments, we had used a methanol carrier solvent, but the polymer used in the 3D-printed GDVN was incompatible with alcohols. Urgent necessity required an alternate solution, and we quickly settled on a 0.2 wt% solution of Dawn dish detergent and water. This solution, affectionately labelled ‘duck sauce’ in honour of the friendly yellow duck featured on the product label, largely solved the issues with both clogging the capillary lines and compatibility with the GDVN. There was no evidence of reduced signal-to-noise values from inclusion of the detergent or contamination in the diffraction patterns. Sample concentration was adjusted to optimize hit rate as measured by spot-finding.

The XFEL beam was tuned to an average per-pulse photon energy of 9.8 keV with a repetition rate of 120 Hz. The per-pulse photon energy was measured using a downstream single-shot spectrometer⁴³. Diffraction patterns were recorded on a Jungfrau 4M detector⁴⁴.

Detector metrology

To synthesize virtual powder patterns from spot-finding results, highly accurate spot positions are required; therefore, the experimental geometry must be measured with high accuracy as well. We achieved this in a two-step process: first, the beam centre and detector distance were determined using powder diffraction rings from silver behenate; second, the internal detector metrology (relative positions of panels) was refined against diffraction from a high-quality protein reference sample.

Silver behenate powder samples were prepared by pressing Kapton tape into a dry powder. The tape was placed at the XFEL interaction point and exposed for ~5 min, after which the data were averaged together and dials.image_viewer was used to adjust the beam centre and detector distance.

For the metrology of the MPCCD detector at SACLAA, protein diffraction data were collected in seven 10-min runs using crystals of Proteinase K from *Engyodontium album* (PDB reference 6k2x). For the detector at LCLS, we used four 5-min runs with crystals of hen egg-white lysozyme (PDB reference 3wun). The refinements were as described previously⁴⁵. For the SACLAA data we found it was necessary to refine individual panel rotations around all three axes. For the LCLS data the detector tilt was refined with all panels coplanar. The final spot position RMS deviations in the protein refinements were 14.1 μm (SACLAA) and 26.6 μm (LCLS).

Synthetic powder patterns from XFEL shots

We use the following approach to construct the powder pattern for unit-cell determination: (1) The DIALS spot-finder is used to harvest Bragg spots from the full dataset. This locates spot centroids with subpixel accuracy and thus eliminates peak broadening generated by sample (crystallite size, strain) and instrumental (axial divergence, detector point spread) factors. (2) The spot positions are converted to d spacings using the per-shot measured XFEL photon energy. (3) The measured d spacings are binned into a histogram that amounts to a sharpened powder diffraction pattern (Fig. 2b). This sharpening of the powder pattern is only possible because we observe a single crystal per shot, which differs from a conventional powder diffraction experiment. Peak resolution is often a limiting factor when attempting to identify a unit cell from powder data, so our technique gives a powerful advantage for cell determination. This technique is implemented as a new CCTBX script, `cctbx.xfel.powder_from_spots`.

Unit-cell determination

Approximately 20 strong peaks were hand-selected from the powder pattern as input for indexing. Trials of GSASIIindex were performed for all Bravais lattices with a starting trial cell volume of 200 \AA^3 . The obtained candidate lattices were ranked by a figure of merit:

$$M_{\text{20}} = \frac{\sum_{i=1}^{20} (d_i - d_{\text{pred}})^2}{\Delta d} M_{\text{20}}$$

where M_{20} is as defined by de Wolff⁴⁶ and quantifies the normalized errors in $1/d^2$ of the peaks selected for the unit-cell search. The summation is over the points in the synthetic powder pattern, with Δd calculated to the nearest peak predicted for the candidate cell. To complete the unit cell determination, we used the best 20 candidate unit cells (10 triclinic and 10 higher symmetry) to index diffraction frames by the `cctbx.small_cell` algorithm. The four unit cells that had the highest indexing rates, labelled in Fig. 3e, can be viewed in Extended Data Table 2. The correct unit cell indexed ~50% more diffraction frames than other candidate cells. Structure solution was attempted for the four candidates with reasonable indexing rates and was only successful for the correct cell. A utility for dispatching GSASIIindex trials and ranking the results is implemented in the new CCTBX script `cctbx.xfel.candidate_cells`.

General methods for data processing

All processing of diffraction data was performed in a DIALS/CCTBX^{35,47} environment, with code and instructions available from: https://github.com/cctbx/cctbx_project/tree/master/xfel/small_cell and links therein.

After unit-cell determination as described above, indexing and integration were performed in cctbx.small_cell_process. An accurate reference geometry was provided for indexing; this was determined by refinement of recently collected diffraction frames from Proteinase K (SACL A data collection) or hen egg-white lysozyme (LCLS data collection). The space group for indexing was a non-centred space group from the previously determined Bravais lattice, that is, space group *C*2 for the *C*-centred monoclinic lattices presented here. Crystal orientations were refined, but unit-cell refinement was disabled. It was necessary to decrease the fractional indexing cut-off from $0.3 \times$ the reciprocal axis length, optimized for protein diffraction, to $0.1 \times$. This change has been included in cctbx.small_cell_process.

Merging was performed in a two-step process with no reference structure using cctbx.xfel.merge, a new version of the cctbx.xfel merging programme²⁹ that is designed for merging large numbers of images. First, the symmetry-equivalent measurements were merged using simple averaging. This dataset was then used as a scaling reference to re-merge the data, scaling each image to the reference set. This two-step approach was previously used to merge lysozyme XFEL diffraction data with accuracy that permitted de novo SAD (single-wavelength anomalous diffraction) phasing⁴⁸. We note that in contrast to typical macromolecular SFX studies, the best results were obtained when we made no attempt to apply a separate resolution cut-off to individual frames on the basis of sparsity of spots on each image. Instead, the overall resolution cut-off was chosen to give $10\text{--}15 \times$ multiplicity for the outer shell in the final dataset. Errors in the merged reflection intensities were estimated using the standard error of the mean of the integrated reflections for each *hkl*, as described by protocol 1 published previously⁴⁸. The results of this second merging trial were usable for structure solution by either olex2.solve¹⁵ or ShelXT¹⁷ as described in the individual refinement sections.

After the structure solution, we repeated the scaling and merging process using the working structure as a scaling reference. This is beneficial because the scaling model is known to be physical (atomic, with positive definite displacement parameters, and so on) in contrast to the previous reference intensities obtained by simple averaging. The scaling models were simplified with isotropic atomic displacement parameters (ADPs) fixed at reasonable values, typically 0.05 \AA^2 for heavy atoms and 0.06 \AA^2 for carbon. At this stage, we typically also tested pseudomerohedral reindexing operators as described in the individual refinement sections. We determined that X-ray absorption would be negligible ($T_{\min} > 0.90$) for crystals in the $3\text{--}10 \mu\text{m}$ range; thus, no absorption correction was considered.

Refinement of mithrene structure

The space group was assigned by standard methods. The presence of a *c* glide plane was indicated by the systematic absences. Space group *C*2/*c* was assigned tentatively because it is $\sim 8\times$ more common than *Cc* for small-molecule structures. Charge-flipping in olex2.solve¹⁵ yielded a successful solution, revealing the three independent atoms in the heavy-atom substructure, one Se atom on a general position and two Ag atoms on twofold special positions. Ag and Se were distinguishable by their peak heights. The heavy atom positions and isotropic ADPs were refined within a 1.2-Å resolution cut-off; this initial refinement converged with a poor R_1 agreement factor of 30.2%. The data were tested for twinning using PLATON⁴⁹ and revealed the twin operator ($-1\ 0\ 0/\ 0\ -1\ 0/\ 1\ 0\ 1$), which was included in the refinement, improving the R_1 factor to 24.3%. We found that the apparent twinning was explained by a pseudo-orthorhombic metric (true: monoclinic C; $a, b, c = 5.94, 7.32, 29.20\ \text{\AA}$; $\beta = 95.4^\circ$; pseudosymmetric: orthorhombic F; $a, b, c = 5.94, 7.32, 58.14\ \text{\AA}$; $\alpha, \beta, \gamma = 90^\circ, 90^\circ, 89.56^\circ$). This metric pseudosymmetry presented an indexing ambiguity with some frames indexed incorrectly, thus simulating pseudo-merohedral twinning of the dataset. We applied the cctbx.xfel.merge tool `modify_reindex_to_reference`, reorienting individual frames by the twin operator to best match the reference structure.

After reindexing and merging with the heavy-atom substructure as a reference, our data were good enough to identify carbon atoms in difference maps (Fig. 3a,b). The C_6H_5 group was located in two positions disordered by rotation around the Se–C bond axis. In further refinement, the two disordered C_6 rings were each constrained to fit an ideal hexagon. Anisotropic ADPs were refined for Ag and Se; a single isotropic ADP was refined for the carbon atoms; hydrogen atoms were placed in calculated positions with riding isotropic ADPs. Finally, we performed one more round of merging with the full SFX-determined structure as a reference for scaling and reindexing. The final refinement converged with R_1 (all) = 13.3% and $R_1(I > 2\sigma(I))$ = 11.4%.

Refinement of thiorene structure

We initially solved the structure in *C*2/*c* using olex2.solve and found that the heavy-atom positions formed a distorted analogue of the mithrene structure. The difference maps did not reveal the phenyl ring, so we tested several positions of a constrained C_6 ring with the aid of 1,2- and 1,3-distance restraints to regularize the geometry. No satisfactory placement was found; in all candidate orientations of the phenyl ring, there were nonbonded H–H clashes or severely distorted geometry around the S atom. Furthermore, there was no ordered model in *C*2/*c* that gave agreement factors better than $R_1 \approx 19\%$ for all data within 1.35-Å resolution. We tested a disordered model for the SC_6H_5 group, which refined to a $\sim 1:1$ occupancy ratio with improved agreement factors, but it seemed unlikely that the entire bonding framework of the material would

be disordered over two independent positions in a statistical ratio. Thus, we finally tested an ordered model in *Cc* with the top and bottom sides of each AgS sheet independent from each other. This immediately improved R_1 to 14% for all data; furthermore, the geometry around S and the nonbonded H–H contacts within the thiorene sheets were all chemically reasonable. We placed hydrogen atoms in calculated positions and temporarily fixed all ADPs at reasonable isotropic values, 0.05 \AA^2 for Ag, S and 0.06 \AA^2 for C. This model was used as a final scaling reference; reindexing was also performed to attempt to resolve the orientation of the polar axis in *Cc*, although it is unlikely that this is resolvable for individual shots with current methods. The final refinement was performed with anisotropic ADPs (stabilized by restraints) for Ag and S and group isotropic ADPs for each of the phenyl rings. The phenyl rings were refined as rigid hexagons and the α -carbons were restrained to a chiral volume of 0. As with mithrene, we detected ‘twinning’ (misindexing) by a twofold rotation around the c^* reciprocal axis, but with a smaller fraction of misindexed frames (<5%). We note that the inversion pseudosymmetry gives larger uncertainties on the refined bonding distances than would otherwise be expected. The Flack parameter was not refined. The final refinement had R_1 (all) = 12.1% and $R_1(I > 2\sigma(I))$ = 10.5%.

Refinement of tethrene structure

After the initial structure solution and isotropic refinement, a difference map revealed the α -C atom of the phenyl group, enabling us to locate the ring as a rigid fragment. After further refinement, the isotropic ADPs were fixed at typical values (0.05 \AA^2 for Ag, Te, and 0.06 \AA^2 for C). The resulting model was used as a reference for another round of scaling and merging using cctbx.xfel.merge to obtain a final dataset. Finally, hydrogen atoms were placed in calculated positions and anisotropic ADPs were refined for the Ag and Te atoms. The geometry of the phenyl ring was stabilized with restraints on C–Te 1,3-distances and on the chiral volume of the α -carbon. The final refinement had R_1 (all) = 13.1% and $R_1(I > 2\sigma(I))$ = 10.6%.

Model bias test

Merging of mithrene diffraction data was performed using several different reference models; the structures were assessed by their root-mean-square deviation (RMSD) agreement (Ag and Se positions only) with the literature single-crystal structure¹⁹ determined at 180 K (Extended Data Fig. 4, Extended Data Table 3). When reindexing and merging was performed with any reasonable model as a reference, the RMSD agreement of heavy-atom positions was excellent, around 0.015 to 0.02 Å. When the data were merged with no reference, either without or with a reindexing step, the RMSD agreement was slightly elevated to 0.025–0.03 Å. We explicitly tested for bias

by ‘sabotaging’ a reference structure with a silver atom displaced by \sim 0.2 Å from its correct position. The resulting final structure had the silver atom displaced by a much smaller distance of 0.028 Å. Thus, although the reference model affects the final refinement, the refined parameters are determined mostly by the diffraction data and only slightly affected by the choice of reference model. The final SFX structure is in good agreement with the literature structure (Extended Data Table 1), with all bonding distances agreeing within 0.04 Å.

Data availability

Crystallographic data for SFX-collected mithrene, thiorene and tethrene structures are available free of charge via the Cambridge Crystallographic Data Centre (CCDC), with deposition numbers 2059768, 2094416 and 2094415, respectively. Raw XFEL SFX data for mithrene, thiorene and tethrene are available free of charge via the Coherent X-ray Imaging Database (CXIDB), accession number 189, <https://www.cxidb.org/id-189.html>.

Code availability

Code for generating powder patterns and interfacing with GSAS-II was implemented in a DIALS/CCTBX environment, with installation instructions available from https://github.com/cctbx/cctbx_project/tree/master/xfel/small_cell and links therein.

References

1. 1.
Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
2. 2.
Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. B* **72**, 171–179 (2016).
3. 3.
Cole, J. C., Wiggin, S. & Stanzione, F. New insights and innovation from a million crystal structures in the Cambridge Structural Database. *Struct. Dyn.* **6**, 054301 (2019).
4. 4.

Kolb, U., Mugnaioli, E. & Gorelik, T. E. Automated electron diffraction tomography – a new tool for nano crystal structure analysis. *Cryst. Res. Technol.* **46**, 542–554 (2011).

5. 5.

Wang, B. et al. A porous cobalt tetraphosphonate metal–organic framework: accurate structure and guest molecule location determined by continuous-rotation electron diffraction. *Chem. Eur. J.* **24**, 17429–17433 (2018).

6. 6.

Jones, C. G. et al. The CryoEM method: MicroED as a powerful tool for small molecule structure determination. *ACS Cent. Sci.* **4**, 1587–1592 (2018).

7. 7.

Samkian, A. E. et al. Elucidation of diverse solid-state packing in a family of electron-deficient expanded helicenes via microcrystal electron diffraction (MicroED). *Angew. Chem. Int. Ed.* **60**, 2493–2499 (2021).

8. 8.

Wang, B., Zou, X. & Smeets, S. Automated serial rotation electron diffraction combined with cluster analysis: an efficient multi-crystal workflow for structure determination. *IUCrJ* **6**, 854–867 (2019).

9. 9.

Smeets, S., Zou, X. & Wan, W. Serial electron crystallography for structure determination and phase analysis of nanocrystalline materials. *J. Appl. Crystallogr.* **51**, 1262–1273 (2018).

10. 10.

Bücker, R. et al. Serial protein crystallography in an electron microscope. *Nat. Commun.* **11**, 996 (2020).

11. 11.

Meng, G., Zou, X. & Huang, Z. Three-dimensional electron diffraction for structural analysis of beam-sensitive metal-organic frameworks. *Crystals* **11**, 263 (2021).

12. 12.

Liang, M. et al. The Coherent X-ray Imaging instrument at the Linac Coherent Light Source. *J. Synchrotron Radiat.* **22**, 514–519 (2015).

13. 13.

Tono, K. et al. Diverse application platform for hard X-ray diffraction in SACLÀ (DAPHNIS): application to serial protein crystallography using an X-ray free-electron laser. *J. Synchrotron Radiat.* **22**, 532–537 (2015).

14. 14.

Brewster, A. S. et al. Indexing amyloid peptide diffraction from serial femtosecond crystallography: new algorithms for sparse patterns. *Acta Crystallogr. D* **71**, 357–366 (2015).

15. 15.

Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. OLEX2: a complete structure solution, refinement, and analysis program. *J. Appl. Cryst.* **42**, 339–341 (2009).

16. 16.

Sheldrick, G. M. Crystal structure refinement with ShelXL. *Acta Crystallogr. C* **71**, 3–8 (2015).

17. 17.

Sheldrick, G. M. SHELXT – integrated space-group and crystal-structure determination. *Acta Crystallogr. A* **71**, 3–8 (2015).

18. 18.

Yao, K. et al. Strongly quantum-confined blue-emitting excitons in chemically configurable multiquantum wells. *ACS Nano* **15**, 4085–4092 (2021).

19. 19.

Cuthbert, H. L., Wallbank, A. I., Taylor, N. J. & Corrigan, J. F. Synthesis and structural characterization of $[\text{Cu}_{20}\text{Se}_4(\mu_3\text{-SePh})_{12}(\text{PPh}_3)_6]$ and $[\text{Ag}(\text{SePh})]_\infty$. *Z. Anorg. Allg. Chem.* **628**, 2483–2488 (2002).

20. 20.

Schriber, E. A. et al. Mithrene is a self-assembling robustly blue luminescent metal–organic chalcogenolate assembly for 2D optoelectronic applications. *ACS Appl. Nano Mater.* **1**, 3498–3508 (2018).

21. 21.

Dance, I. G., Fisher, K. J., Banda, R. M. H. & Scudder, M. L. Layered structure of crystalline compounds AgSR. *Inorg. Chem.* **30**, 183–187 (1991).

22. 22.

Ozaki, Y. et al. Automated crystal structure analysis based on blackbox optimisation. *npj Comput. Mater.* **6**, 75 (2020).

23. 23.

Jaramillo, D. E. et al. Selective nitrogen adsorption via backbonding in a metal–organic framework with exposed vanadium sites. *Nat. Mater.* **19**, 517–521 (2020).

24. 24.

Coelho, A. A. TOPAS and TOPAS-Academic: an optimization program integrating computer algebra and crystallographic objects written in C++. *J. Appl. Crystallogr.* **51**, 210–218 (2018).

25. 25.

David, W. I. F. et al. DASH: a program for crystal structure determination from powder diffraction data. *J. Appl. Crystallogr.* **39**, 910–915 (2006).

26. 26.

Christensen, J. et al. Radiation damage in small-molecule crystallography: fact not fiction. *IUCrJ* **6**, 703–713 (2019).

27. 27.

Weierstall, U. Liquid sample delivery techniques for serial femtosecond crystallography. *Phil. Trans. R. Soc. B* **369**, 20130337 (2014).

28. 28.

Wiedorn, M. O. et al. Megahertz serial crystallography. *Nat. Commun.* **9**, 4025 (2018).

29. 29.

Hattne, J. et al. Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers. *Nat. Methods* **11**, 545–548 (2014); correction **12**, 692 (2015).

30. 30.

Chapman, H. N. et al. Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011).

31. 31.

Dejoie, C. & Tamura, N. Pattern-matching indexing of Laue and monochromatic serial crystallography data for applications in materials science. *J. Appl. Crystallogr.* **53**, 824–836 (2020).

32. 32.

Dejoie, C. et al. Serial snapshot crystallography for materials science with SwissFEL. *IUCrJ* **2**, 361–370 (2015).

33. 33.

Toby, B. H. & Von Dreele, R. B. GSAS-II: the genesis of a modern open-source all purpose crystallography software package. *J. Appl. Crystallogr.* **46**, 544–549 (2013).

34. 34.

Coelho, A. A. Indexing of powder diffraction patterns by iterative use of singular value decomposition. *J. Appl. Cryst.* **36**, 86–95 (2003).

35. 35.

Winter, G. et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D* **74**, 85–97 (2018).

36. 36.

Maserati, L. et al. Anisotropic 2D excitons unveiled in organic–inorganic quantum wells. *Mater. Horiz.* **8**, 197–208 (2021).

37. 37.

Schmidbaur, H. & Schier, A. Argentophilic interactions. *Angew. Chem. Int. Ed.* **54**, 746–784 (2015).

38. 38.

Paritmongkol, W. Size and quality enhancement of 2D semiconducting metal–organic chalcogenolates by amine addition. *J. Am. Chem. Soc.* **143**, 20256–20263 (2021).

39. 39.

Trang, B. et al. Tarnishing silver metal into mithrene. *J. Am. Chem. Soc.* **140**, 13892–13903 (2018).

40. 40.

Chatterjee, R. et al. XANES and EXAFS of dilute solutions of transition metals at XFELs. *J. Synchrotron Rad.* **26**, 1716–1724 (2019).

41. 41.

Kameshima, T. et al. Development of an X-ray pixel detector with multi-port charge-coupled device for X-ray free-electron laser experiments. *Rev. Sci. Instrum.* **85**, 033110 (2014).

42. 42.

DePonte, D. P. et al. Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D* **41**, 195505 (2008).

43. 43.

Zhu, D. et al. A single-shot transmissive spectrometer for hard X-ray free electron lasers. *Appl. Phys. Lett.* **101**, 034103 (2012).

44. 44.

Leonarski, F. et al. Fast and accurate data collection for macromolecular crystallography using the JUNGFRAU detector. *Nat. Methods* **15**, 799–804

(2018).

45. 45.

Brewster, A. S. et al. Improving signal strength in serial crystallography with DIALS geometry refinement. *Acta Crystallogr. D* **74**, 877–894 (2018).

46. 46.

de Wolff, P. M. The definition of the indexing figure of merit M_{20} . *J. Appl. Crystallogr.* **5**, 243 (1972).

47. 47.

Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W. & Adams, P. D. The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework. *J. Appl. Crystallogr.* **35**, 126–136 (2002).

48. 48.

Brewster, A. S. et al. SAD phasing of XFEL data depends critically on the error model. *Acta Crystallogr. D* **75**, 959–968 (2019).

49. 49.

Spek, A. L. PLATON, an integrated tool for the analysis of the results of a single crystal structure determination. *Acta Crystallogr. A* **46**, 34–34 (1990).

Acknowledgements

We thank the SACLA and LCLS technical staff for their invaluable support. We are also grateful for the computational support from the SACLA HPC system. XFEL experiments were carried out at SACLA with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (proposal number 2019A8074). Research was supported by the NIH grant GM117126 to N.K.S. D.J.R. was supported by DOE-BER IDAT DE-AC02-05CH11231 and NIGMS ALS-ENABLE (P30 GM124169 and S10OD018483). D.M., J.P.B. and N.K.S. were supported by the Exascale Computing Project (grant 17-SC-20-SC), a collaborative effort of the Department of Energy (DOE) Office of Science and the National Nuclear Security Administration. E.A.S., A.S.B. and J.N.H. were supported by DOE ICDI grant DE-SC0022215. C.P.-R. and D.S.-I. were supported by PAPIIT IN216020 and CONACYT's CB-A1-S-8729. P.J.S. and K.Y. acknowledge support by Programmable Quantum Materials, an Energy Frontier Research Center funded by the US Department of Energy (DOE), Office of

Science, Basic Energy Sciences (BES), under award DE-SC0019443. Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the US DOE, Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. Work was performed at the Molecular Foundry and was supported by the Office of Science, Office of Basic Energy Sciences, of the US DOE under contract no. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a US DOE, Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under contract no. DE-AC02-05CH11231. B. H. Toby is acknowledged for assistance with GSAS-II.

Author information

Author notes

1. These authors contributed equally: Elyse A. Schriber, Daniel W. Paley

Affiliations

1. Institute of Materials Science, University of Connecticut, Storrs, CT, USA

Elyse A. Schriber, Ryan P. Kelly & J. Nathan Hohman

2. Department of Chemistry, University of Connecticut, Storrs, CT, USA

Elyse A. Schriber, Ryan P. Kelly & J. Nathan Hohman

3. Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Daniel W. Paley, Robert Bolotovsky, Daniel J. Rosenberg, Derek Mendez, Asmit Bhowmick, Nicholas K. Sauter & Aaron S. Brewster

4. Graduate Group in Biophysics, University of California, Berkeley, CA, USA

Daniel J. Rosenberg

5. Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Raymond G. Sierra, Andrew Aquila, Frédéric Poitevin, Mark Hunter & Brandon Hayes

6. National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Johannes P. Blaschke & Derek C. Popple

7. College of Chemistry, University of California, Berkeley, Berkeley, CA, USA

Derek C. Popple

8. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA

Matthew Yeung

9. Instituto de Investigaciones en Materiales, Universidad Nacional Autónoma de México, Coyoacán, Mexico

Carina Pareja-Rivera & Diego Solis-Ibarra

10. Department of Physics, Arizona State University, Tempe, AZ, USA

Stella Lisova

11. SPring-8, Japan Synchrotron Radiation Research Institute, Sayo, Japan

Kensuke Tono & Shigeki Owada

12. RIKEN SPring-8 Center, Sayo, Japan

Michihiro Sugahara

13. The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Tevye Kuykendall

14. Department of Mechanical Engineering, Columbia University, New York, NY, USA

Kaiyuan Yao & P. James Schuck

Contributions

smSFX experiments were conceived and designed by E.A.S., D.W.P., A.S.B. and J.N.H. and were executed and analysed by E.A.S., D.W.P., R.B., D.M., J.P.B., A.B., D.J.R., R.G.S., M.H., A.A., B.H., C.P.-R., K.T., S.L., M.S., S.O., K.Y., P.J.S., T.K., D.S.-I., N.K.S., A.S.B. and J.N.H. Synthetic work was performed by E.A.S., R.P.K., D.C.P., M.Y. and J.N.H. The manuscript was written through contributions from all authors.

Corresponding authors

Correspondence to [Nicholas K. Sauter](#), [Aaron S. Brewster](#) or [J. Nathan Hohman](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Tim Grüne, Lukáš Palatinus and Paul Raithby 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 Single-shot frame from the Proteinase K dataset that was collected at SACLA for detector metrology on the MPCCD detector.**](#)

Indexed reflections are in red.

[**Extended Data Fig. 2 Single-shot frame from the mithrene dataset.**](#)

Indexed reflections are boxed in the image and shown in the insets.

[Extended Data Fig. 3 Indexing statistics using the XFEL-determined room-temperature mithrene unit cell.](#)

a, Each spot from 6,577 mithrene crystals is shown on the detector simultaneously, coloured by the magnitude of the difference (in μm) between predicted and observed spot locations on the detector. Using the literature value cryogenic unit cell¹⁹ to index the whole XFEL dataset produced an overall RMSD between predicted and observed reflections of $123.3\ \mu\text{m}$. **b**, Using the XFEL determined the room-temperature unit cell improved indexing accuracy, with an overall RMSD of $110.6\ \mu\text{m}$ and increased the number of indexed high-resolution reflections, seen in the insets of the bottom corner panel on the detector.

[Extended Data Fig. 4 Iterative reindeindexing and merging with reference model bias test results.](#)

R_1 is calculated for all data. RMSD is the deviation of heavy atom (Ag, Se) positions in angstroms, from the reference structure in ref. ¹⁹.

Extended Data Table 1 Interatomic distances in mithrene measured by SFX and SCXRD¹⁹

Extended Data Table 2 Candidate mithrene unit-cell dimensions, where cell 2 is the correct room-temperature unit cell

Extended Data Table 3 Descriptions of merging protocols for mithrene structure determination and model bias test

Supplementary information

[Peer Review File](#)

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the

permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

[Reprints and Permissions](#)

About this article

Cite this article

Schriber, E.A., Paley, D.W., Bolotovsky, R. *et al.* Chemical crystallography by serial femtosecond X-ray diffraction. *Nature* **601**, 360–365 (2022).
<https://doi.org/10.1038/s41586-021-04218-3>

- Received: 30 July 2021
- Accepted: 07 November 2021
- Published: 19 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04218-3>

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/s41586-021-04218-3>

- Article
- [Published: 19 January 2022](#)

Enantiomer-dependent immunological response to chiral nanoparticles

- [Liguang Xu](#)^{1,2} ✉ na1,
- [Xiuxiu Wang](#)^{1,2} ✉ na1,
- [Weiwei Wang](#)^{1,2},
- [Maozhong Sun](#)^{1,2},
- [Won Jin Choi](#) [ORCID: orcid.org/0000-0001-6303-8899](#)³,
- [Ji-Young Kim](#) [ORCID: orcid.org/0000-0001-5631-957X](#)⁴,
- [Changlong Hao](#)^{1,2},
- [Si Li](#)⁵,
- [Aihua Qu](#)^{1,2},
- [Meiru Lu](#)^{1,2},
- [Xiaoling Wu](#)^{1,2},
- [Felippe M. Colombari](#) [ORCID: orcid.org/0000-0002-5431-1301](#)⁶,
- [Weverson R. Gomes](#) [ORCID: orcid.org/0000-0002-9456-9534](#)⁷,
- [Asdrubal L. Blanco](#)⁷,
- [Andre F. de Moura](#) [ORCID: orcid.org/0000-0003-1974-4803](#)⁷,
- [Xiao Guo](#)^{1,2},
- [Hua Kuang](#) ^{1,2,5,8},
- [Nicholas A. Kotov](#) [ORCID: orcid.org/0000-0002-6864-5804](#)^{3,4} &
- [Chuanlai Xu](#) [ORCID: orcid.org/0000-0002-5639-7102](#)^{1,2}

Nature volume **601**, pages 366–373 (2022)

- 4485 Accesses

- 1 Citations
- 33 Altmetric
- [Metrics details](#)

Subjects

- [Applied immunology](#)
- [Nanomedicine](#)

Abstract

Chirality is a unifying structural metric of biological and abiological forms of matter. Over the past decade, considerable clarity has been achieved in understanding the chemistry and physics of chiral inorganic nanoparticles^{1,2,3,4}; however, little is known about their effects on complex biochemical networks^{5,6}. Intermolecular interactions of biological molecules and inorganic nanoparticles show some commonalities^{7,8,9}, but these structures differ in scale, in geometry and in the dynamics of chiral shapes, which can both impede and strengthen their mirror-asymmetric complexes. Here we show that achiral and left- and right-handed gold biomimetic nanoparticles show different in vitro and in vivo immune responses. We use irradiation with circularly polarized light (CPL) to synthesize nanoparticles with controllable nanometre-scale chirality and optical anisotropy factors (*g*-factors) of up to 0.4. We find that binding of nanoparticles to two proteins from the family of adhesion G-protein-coupled receptors (AGPCRs)—namely cluster-of-differentiation 97 (CD97) and epidermal-growth-factor-like-module receptor 1 (EMR1)—results in the opening of mechanosensitive potassium-efflux channels, the production of immune signalling complexes known as inflammasomes, and the maturation of mouse bone-marrow-derived dendritic cells. Both in vivo and in vitro immune responses depend monotonically on the *g*-factors of the nanoparticles, indicating that nanoscale chirality can be used to regulate the maturation of immune cells. Finally, left-handed nanoparticles show substantially higher (1,258-fold) efficiency compared with their right-

handed counterparts as adjuvants for vaccination against the H9N2 influenza virus, opening a path to the use of nanoscale chirality in immunology.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Morphology and spectroscopy of photosynthesized chiral nanoparticles (NPs).

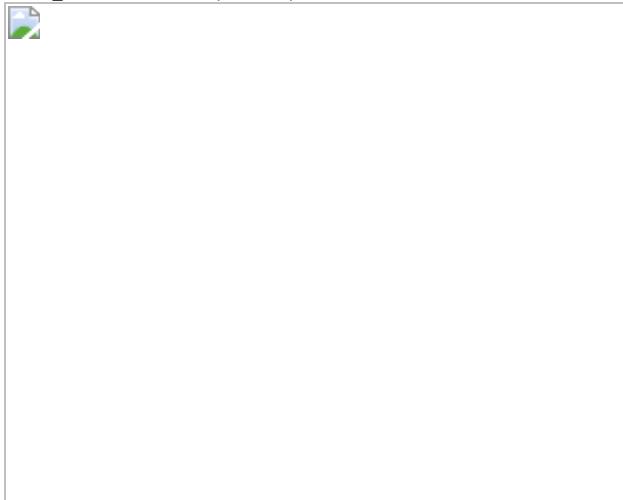


Fig. 2: Quantification of electromagnetic fields and chirality measures for photosynthesized nanoparticles.

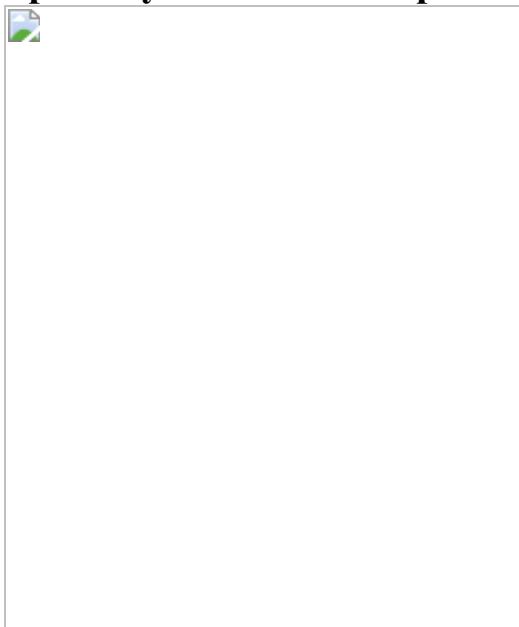


Fig. 3: Nanoparticle-mediated immune responses.

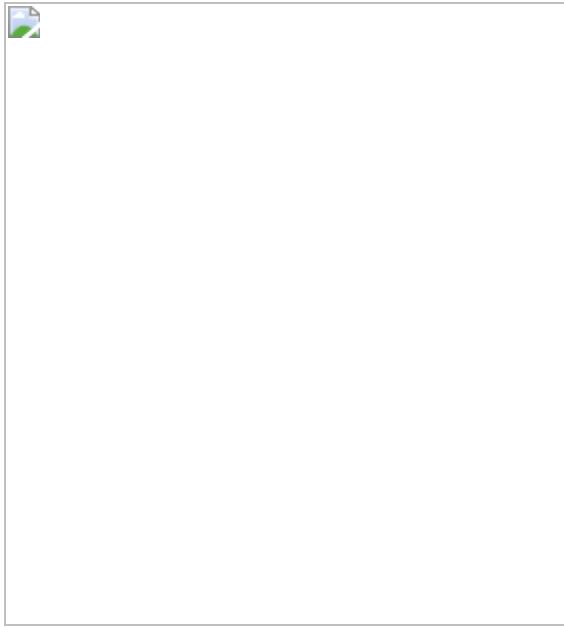
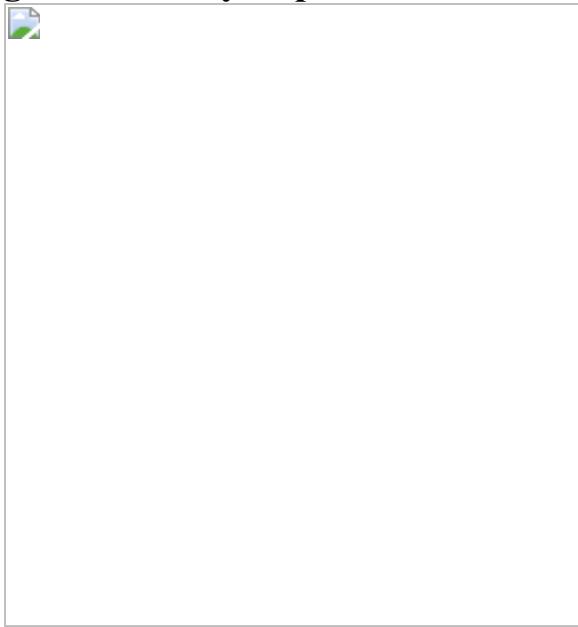


Fig. 4: Chirality-dependent intracellular intake of BMDCs.



Data availability

Source data for Figs. 3, 4 and Extended Data Figs. 1, 2 are provided with this paper. The data supporting the findings of this study are available within the paper and its [Supplementary Information](#) files. [Source data](#) are provided with this paper.

References

1. 1.

Ma, W. et al. Chiral inorganic nanostructures. *Chem. Rev.* **117**, 8041–8093 (2017).

2. 2.

Copeland, L. O. & McDonald, M. B. in *Principles of Seed Science and Technology* 59–110 (Springer, 1999).

3. 3.

Zhang, Q. et al. Unraveling the origin of chirality from plasmonic nanoparticle-protein complexes. *Science* **365**, 1475–1478 (2019).

4. 4.

Guerrero-Martínez, A., Alonso-Gómez, J. L., Auguié, B., Cid, M. M. & Liz-Marzán, L. M. From individual to collective chirality in metal nanoparticles. *Nano Today* **6**, 381–400 (2011).

5. 5.

Kuznetsova, V. A. et al. Enantioselective cytotoxicity of ZnS:Mn quantum dots in A549 cells. *Chirality* **29**, 403–408 (2017).

6. 6.

Sun, M. et al. Site-selective photoinduced cleavage and profiling of DNA by chiral semiconductor nanoparticles. *Nat. Chem.* **10**, 821–830 (2018).

7. 7.

Kotov, N. A. Inorganic nanoparticles as protein mimics. *Science* **330**, 188–189 (2010).

8. 8.

Cagno, V. et al. Broad-spectrum non-toxic antiviral nanoparticles with a virucidal inhibition mechanism. *Nat. Mater.* **17**, 195–203 (2018).

9. 9.

Wang, D. et al. Engineering nanoparticles to locally activate T cells in the tumor microenvironment. *Sci. Immunol.* **4**, eaau6584 (2019).

10. 10.

Gérard, V. A. et al. Plasmon-induced CD response of oligonucleotide-conjugated metal nanoparticles. *Chem. Commun.* **47**, 7383 (2011).

11. 11.

Yeom, J. et al. Chiromagnetic nanoparticles and gels. *Science* **359**, 309–314 (2018).

12. 12.

Ma, W. et al. Attomolar DNA detection with chiral nanorod assemblies. *Nat. Commun.* **4**, 2689 (2013).

13. 13.

Zheng, G. et al. Tuning the morphology and chiroptical properties of discrete gold nanorods with amino acids. *Angew. Chem. Int. Edn* **57**, 16452–16457 (2018).

14. 14.

Chen, W. et al. Nanoparticle Superstructures Made by Polymerase Chain Reaction: Collective Interactions of Nanoparticles and a New Principle for Chiral Materials. *Nano Lett.*, **9**, 2153–2159 (2009).

15. 15.

Singh, G. et al. Self-assembly of magnetite nanocubes into helical superstructures. *Science* **345**, 1149–1153 (2014).

16. 16.

Molotsky, T., Tamarin, T., Ben Moshe, A., Markovich, G. & Kotlyar, A. B. Synthesis of chiral silver clusters on a DNA template. *J. Phys. Chem. C* **114**, 15951–15954 (2010).

17. 17.

Im, S. W. et al. Chiral surface and geometry of metal nanocrystals. *Adv. Mater.* **32**, 1905758 (2020).

18. 18.

Wang, J. et al. Physical activation of innate immunity by spiky particles. *Nat. Nanotechnol.* **13**, 1078–1086 (2018).

19. 19.

Geva, M., Frolow, F., Eisenstein, M. & Addadi, L. Antibody recognition of chiral surfaces. Enantiomorphous crystals of leucine-leucine-tyrosine. *J. Am. Chem. Soc.* **125**, 696–704 (2003).

20. 20.

Walkey, C. D., Olsen, J. B., Guo, H., Emili, A. & Chan, W. C. W. Nanoparticle size and surface chemistry determine serum protein adsorption and macrophage uptake. *J. Am. Chem. Soc.* **134**, 2139–2147 (2012).

21. 21.

del Pino, P. et al. Protein corona formation around nanoparticles—from the past to the future. *Mater. Horiz.* **1**, 301–313 (2014).

22. 22.

Wang, X. et al. Chiral surface of nanoparticles determines the orientation of adsorbed transferrin and its interaction with receptors. *ACS Nano* **11**, 4606–4616 (2017).

23. 23.

Kim, J.-Y. et al. Assembly of gold nanoparticles into chiral superstructures driven by circularly polarized light. *J. Am. Chem. Soc.* **141**, 11739–11744 (2019).

24. 24.

Yeom, J. et al. Chiral templating of self-assembling nanostructures by circularly polarized light. *Nat. Mater.* **14**, 66–72 (2015).

25. 25.

Ou, Z., Wang, Z., Luo, B., Luijten, E. & Chen, Q. Kinetic pathways of crystallization at the nanoscale. *Nat. Mater.* **19**, 450–455 (2020).

26. 26.

Karst, J. et al. Chiral scatterometry on chemically synthesized single plasmonic nanoparticles. *ACS Nano* **13**, 8659–8668 (2019).

27. 27.

González-Rubio, G. et al. Femtosecond laser reshaping yields gold nanorods with ultranarrow surface plasmon resonances. *Science* **358**, 640–644 (2017).

28. 28.

Saito, K. & Tatsuma, T. Chiral plasmonic nanostructures fabricated by circularly polarized light. *Nano Lett.* **18**, 3209–3212 (2018).

29. 29.

Lee, H.-E. et al. Amino-acid- and peptide-directed synthesis of chiral plasmonic gold nanoparticles. *Nature* **556**, 360–365 (2018).

30. 30.

Zhang, Q. et al. Neutrophil membrane-coated nanoparticles inhibit synovial inflammation and alleviate joint damage in inflammatory arthritis. *Nat. Nanotechnol.* **13**, 1182–1190 (2018).

31. 31.

Pelliccia, M. et al. Additives for vaccine storage to improve thermal stability of adenoviruses from hours to months. *Nat. Commun.* **7**, 13520 (2016).

32. 32.

Xia, Y. et al. Exploiting the pliability and lateral mobility of Pickering emulsion for enhanced vaccination. *Nat. Mater.* **17**, 187–194 (2018).

33. 33.

Langenhan, T., Aust, G. & Hamann, J. Sticky signaling—adhesion class G protein-coupled receptors take the stage. *Sci. Signal.* **6**, re3 (2013).

34. 34.

Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat. Rev. Mol. Cell Biol.* **9**, 60–71 (2008).

35. 35.

Ferguson, S. M. & De Camilli, P. Dynamin, a membrane-remodelling GTPase. *Nat. Rev. Mol. Cell Biol.* **13**, 75–88 (2012).

36. 36.

Richards, D. M. & Endres, R. G. Target shape dependence in a simple model of receptor-mediated endocytosis and phagocytosis. *Proc. Natl Acad. Sci. USA* **113**, 6113–6118 (2016).

37. 37.

Mahmoudi, M., Azadmanesh, K., Shokrgozar, M. A., Journeay, W. S. & Laurent, S. Effect of nanoparticles on the cell life cycle. *Chem. Rev.* **111**, 3407–3432 (2011).

38. 38.

Murray, P. J. & Wynn, T. A. Protective and pathogenic functions of macrophage subsets. *Nat. Rev. Immunol.* **11**, 723–737 (2011).

39. 39.

Ohta, S., Glancy, D. & Chan, W. C. W. DNA-controlled dynamic colloidal nanoparticle systems for mediating cellular interaction. *Science* **351**, 841–845 (2016).

40. 40.

Naur, P. et al. Ionotropic glutamate-like receptor 2 binds D-serine and glycine. *Proc. Natl Acad. Sci. USA* **104**, 14116–14121 (2007).

41. 41.

Cobb, M. H. & Ross, E. M. in *Mol. Biol. Cell* 6th edn (eds Alberts, B. et al.) 589–643 (Garland, 2002).

42. 42.

Muñoz-Planillo, R. et al. K⁺ efflux is the common trigger of NLRP3 inflammasome activation by bacterial toxins and particulate matter. *Immunity* **38**, 1142–1153 (2013).

43. 43.

Kefauver, J. M., Ward, A. B. & Patapoutian, A. Discoveries in structure and physiology of mechanically activated ion channels. *Nature* **587**, 567–576 (2020).

44. 44.

Ranade, S. S., Syeda, R. & Patapoutian, A. Mechanically activated ion channels. *Neuron* **87**, 1162–1179 (2015).

45. 45.

Galic, M. et al. External push and internal pull forces recruit curvature-sensing N-BAR domain proteins to the plasma membrane. *Nat. Cell Biol.* **14**, 874–881 (2012).

46. 46.

Chen, L. et al. High-yield seedless synthesis of triangular gold nanoplates through oxidative etching. *Nano Lett.* **14**, 7201–7206 (2014).

47. 47.

Johnson, P. B. & Christy, R. W. Optical constants of the noble metals. *Phys. Rev. B* **6**, 4370–4379 (1972).

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (21925402, 32071400, 21977038, 92156003), Natural Science Foundation of Jiangsu Province (BK20212014), and Young Elite Scientist Sponsorship Program by China Association for Science and Technology (CAST) (2019QNRC001). N.A.K thanks the US National Science Foundation (NSF) (grants 1463474 and 1566460) for support. A.F.M. thanks the Brazilian Ministério da Educação (MEC)/Tutorial Education Programme (PET) for a fellowship and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for a research fellowship (311353/2019-3). We thank the Brazilian funding agencies

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), CNPq, and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (processes 2012/15147-4 and 2013/07296-2) for financial support; and the high-performance computer (HPC) resources provided by the SDumont supercomputer at the National Laboratory for Scientific Computing (Laboratório Nacional de Computação Científica (LNCC)/Ministério da Ciência, Tecnologia e Inovações (MCTI), Brazil; <http://sdumont.lncc.br>) and by the Cloud@UFSCar.

Author information

Author notes

1. These authors contributed equally: Liguang Xu, Xiuxiu Wang

Affiliations

1. State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu, China

Liguang Xu, Xiuxiu Wang, Weiwei Wang, Maozhong Sun, Changlong Hao, Aihua Qu, Meiru Lu, Xiaoling Wu, Xiao Guo, Hua Kuang & Chuanlai Xu

2. International Joint Research Laboratory for Biointerface and Biodetection, School of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu, China

Liguang Xu, Xiuxiu Wang, Weiwei Wang, Maozhong Sun, Changlong Hao, Aihua Qu, Meiru Lu, Xiaoling Wu, Xiao Guo, Hua Kuang & Chuanlai Xu

3. Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI, USA

Won Jin Choi & Nicholas A. Kotov

4. Department of Chemical Engineering, Biointerface Institute,
University of Michigan, Ann Arbor, MI, USA

Ji-Young Kim & Nicholas A. Kotov

5. The Key Laboratory of Synthetic and Biological Colloids, Ministry of
Education, School of Chemical and Material Engineering, Jiangnan
University, Wuxi, Jiangsu, China

Si Li & Hua Kuang

6. Brazilian Nanotechnology National Laboratory, Brazilian Center for
Research in Energy and Materials, Campinas, São Paulo, Brazil

Felippe M. Colombari

7. Department of Chemistry, Federal University of São Carlos, São
Carlos, São Paulo, Brazil

Weverson R. Gomes, Asdrubal L. Blanco & Andre F. de Moura

8. Science Center for Future Foods, Jiangnan University, Wuxi, Jiangsu,
China

Hua Kuang

Contributions

H.K., N.A.K. and C.X. conceived the project and planned the experiments.
L.X., X. Wang, C.H., S.L. and X. Wu fabricated the chiral nanoparticles.
L.X., W.W., M.S., A.Q., M.L. and X.G. carried out immunological
experiments. X. Wang and X.G. measured the affinity constant between
chiral nanoparticles and receptors. L.X., M.S., A.Q. and M.L. carried out
the inflammasome experiments. W.J.C. carried out electromagnetic
modelling and simulation. J.-Y.K. calculated the chirality measures. F.M.C.,
W.R.G., A.L.B. and A.F.D. carried out electrodynamics calculations. L.X.,
H.K., N.A.K. and C.X. conceptualized the work, wrote the manuscript, and

compiled figures, with discussion of results and feedback on the manuscript from all authors.

Corresponding authors

Correspondence to [Hua Kuang](#), [Nicholas A. Kotov](#) or [Chuanlai Xu](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Jacques Neefjes, Luke O'Neill and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Additional information

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

Extended data figures and tables

[Extended Data Fig. 1 Chiral nanoparticles are taken up by human BMDCs and activate inflammasomes.](#)

a, Flow-cytometry data for human BMDCs after being treated with PBS, anti-EMR1 antibody ($30 \mu\text{g ml}^{-1}$, blocking EMR1), anti-CD97 antibody ($20 \mu\text{g ml}^{-1}$, blocking CD97), or both anti-EMR1 antibody ($30 \mu\text{g ml}^{-1}$) and anti-CD97 antibody ($20 \mu\text{g ml}^{-1}$) (blocking both CD97 and EMR1) and then incubated with l-P⁺ NP(2 nM) or d-P⁻ NP (2 nM) for 8 h. **b**, Confocal

imaging of human BMDCs incubated with 2 µg ml⁻¹ MPL, 20 µg ml⁻¹ OVA and 2 nM l-P⁺ NP with various incubation times up to 4 h. Blue, DAPI; red, CD97–Cy5; green: l-P⁺ NP–Cy3. Scale bar, 10 µm. **c**, Confocal imaging of NLRP3 inflammasome activation in mouse BMDCs after incubation with PBS, MPL plus OVA, l-P⁺ NP + MPL + OVA, l-P⁺ NP + MPL + OVA + MCC950 (NLRP3 inhibitor), l-P⁺ NP + MPL + OVA + amiodarone (K⁺-channel inhibitor), l-P⁺ NP + MPL + OVA + KCl (K⁺-efflux inhibitor), l-P⁺ NP + MPL + OVA + dynasore (dynamin inhibitor), l-P⁺ NP + MPL + OVA + chlorpromazine (clathrin inhibitor), l-P⁺ NP + MPL + OVA + CA-074-Me (cathepsin B inhibitor), l-P⁺ NP + MPL + OVA + cytochalasin D (phagocytosis inhibitor), l-P⁺ NP + MPL + OVA + NAC (inhibitor of reactive oxygen species, ROS) and l-P⁺ NP + MPL + OVA + nocodazole (microtubule inhibitor) for 12 h. Blue, DAPI; red, caspase-1; green, NLRP3. Scale bar, 20 µm. Data are mean ± s.d. ($n = 5$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, analysed by Student's *t*-test.

[Source data](#)

[Extended Data Fig. 2 Chirality-dependent efficiency of vaccination in mice.](#)

a, IL-1 β concentration in the culture medium of mouse BMDCs after incubation with chiral nanoparticles of different *g*-factors, measured by ELISA. **b**, Expression of NLRP3 in wild-type mice after treatment with chiral nanoparticles of different *g*-factors, detected by flow cytometry. **c–f**, Influenza vaccination. C57BL/6 mice ($n = 5$) were immunized with H9N2 influenza vaccine and the indicated adjuvants, including MPL, alum + MPL, d-P⁻ NP + MPL, l-P⁺ NP + MPL, NS-d-CYP + MPL, or NS-l-CYP + MPL. **c**, The serum of the mice was collected to measure vaccine-specific antibody titres. **d–f**, IFN- γ -secreting CD8 $^{+}$ T cells (**d**), IFN- γ -secreting CD4 $^{+}$ T cells (**e**) and IL-4-secreting CD4 $^{+}$ T cells (**f**) in the spleen were measured by flow cytometry 7 days after immunization. Data are means ± s.d. ($n = 5$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, analysed by Student's *t*-test.

[Source data](#)

Supplementary information

[Supplementary Information](#)

This file contains Supplementary Methods; Supplementary Tables 1–3; Supplementary Figures 1–30 and Supplementary References

[Reporting Summary](#)

[Peer Review File](#)

Source data

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

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

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

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Xu, L., Wang, X., Wang, W. *et al.* Enantiomer-dependent immunological response to chiral nanoparticles. *Nature* **601**, 366–373 (2022).
<https://doi.org/10.1038/s41586-021-04243-2>

- Received: 04 March 2020
- Accepted: 15 November 2021
- Published: 19 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04243-2>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [Nanoparticle asymmetry shapes an immune response](#)
 - Alexander Hooftman
 - Luke A. J. O'Neill

Nature (2022)

[Why mutation is not as random as we thought](#)

- Shamini Bundell
- Benjamin Thompson

Nature Podcast 19 Jan 2022

Nanoparticle asymmetry shapes an immune response

- Alexander Hooftman
- Luke A. J. O'Neill

News & Views 19 Jan 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04243-2>

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

- Article
- [Published: 19 January 2022](#)

Historical glacier change on Svalbard predicts doubling of mass loss by 2100

- [Emily C. Geyman](#) ¹nAff4,
- [Ward J. J. van Pelt](#) [ORCID: orcid.org/0000-0003-4839-7900](#)²,
- [Adam C. Maloof](#) [ORCID: orcid.org/0000-0003-0032-6628](#)³,
- [Harald Faste Aas](#)¹ &
- [Jack Kohler](#)¹

Nature volume **601**, pages 374–379 (2022)

- 1424 Accesses
- 74 Altmetric
- [Metrics details](#)

Subjects

- [Cryospheric science](#)
- [Projection and prediction](#)

Abstract

The melting of glaciers and ice caps accounts for about one-third of current sea-level rise^{1,2,3}, exceeding the mass loss from the more voluminous Greenland or Antarctic Ice Sheets^{3,4}. The Arctic archipelago of Svalbard,

which hosts spatial climate gradients that are larger than the expected temporal climate shifts over the next century^{5,6}, is a natural laboratory to constrain the climate sensitivity of glaciers and predict their response to future warming. Here we link historical and modern glacier observations to predict that twenty-first century glacier thinning rates will more than double those from 1936 to 2010. Making use of an archive of historical aerial imagery⁷ from 1936 and 1938, we use structure-from-motion photogrammetry to reconstruct the three-dimensional geometry of 1,594 glaciers across Svalbard. We compare these reconstructions to modern ice elevation data to derive the spatial pattern of mass balance over a more than 70-year timespan, enabling us to see through the noise of annual and decadal variability to quantify how variables such as temperature and precipitation control ice loss. We find a robust temperature dependence of melt rates, whereby a 1 °C rise in mean summer temperature corresponds to a decrease in area-normalized mass balance of -0.28 m yr^{-1} of water equivalent. Finally, we design a space-for-time substitution⁸ to combine our historical glacier observations with climate projections and make first-order predictions of twenty-first century glacier change across Svalbard.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Three-dimensional models of Svalbard glaciers.

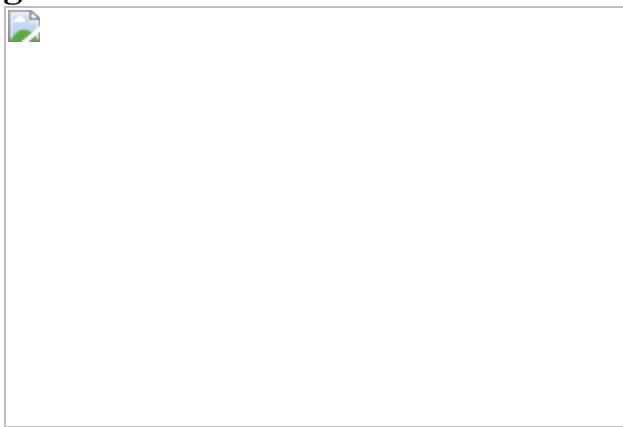


Fig. 2: Elevation change from 1936 to approximately 2010.



Fig. 3: Temperature control on glacier mass balance.

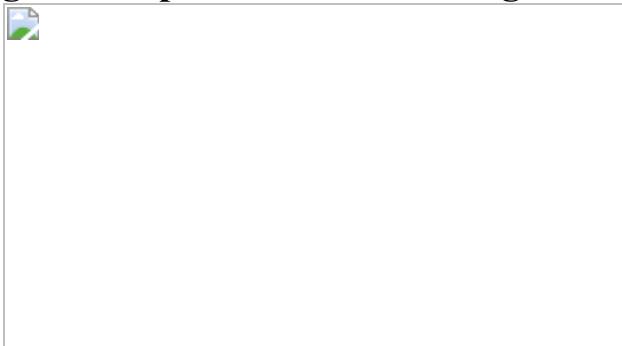
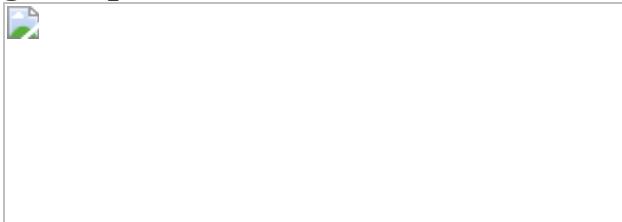


Fig. 4: Space-for-time substitution.



Data availability

The 1936/1938 Svalbard glacier inventory presented here consists of raster DEMs and orthophotos (5 m resolution), and vector outlines of glacier extents (Extended Data Fig. 3). All data are publicly available on the NPI website (<https://www.doi.org/10.21334/npolar.2021.f6afca5c>) and on Zenodo (<https://doi.org/10.5281/zenodo.5806388>). In these repositories, we also provide the raw (unprocessed) 3D point clouds as .laz files and a

spreadsheet (.xlsx file) containing glacier-by-glacier estimates of area, volume, hypsometry, $\Delta h/\Delta t$, bed slope, DEM uncertainty and climate fields (mean annual temperature, mean summer temperature, PDDs, precipitation as snow and total precipitation). The original 1936/1938 aerial images and their locations can be viewed at <https://toposvalbard.npolar.no/>. The 5 m regional DEMs from the 2008–2012 survey³³ are available as .tif files from https://publicdatasets.data.npolar.no/kartdata/S0_Terrengmodell/Delmodell/ and the associated 50 cm orthophotomosaic is available as a WMTS layer from <https://geodata.npolar.no/#basemap-data>.

Code availability

The code developed to analyse the 1936–2010 mass balance data and implement the space-for-time substitution is available on Zenodo (<https://doi.org/10.5281/zenodo.5643856>).

References

1. 1.
Meier, M. F. et al. Glaciers dominate eustatic sea-level rise in the 21st century. *Science* **317**, 1064–1067 (2007).
2. 2.
Gardner, A. S. et al. A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009. *Science* **340**, 852–857 (2013).
3. 3.
Zemp, M. et al. Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568**, 382–386 (2019).
4. 4.
Hugonnet, R. et al. Accelerated global glacier mass loss in the early twenty-first century. *Nature* **592**, 726–731 (2021).

5. 5.

van Pelt, W. et al. A long-term dataset of climatic mass balance, snow conditions, and runoff in Svalbard (1957–2018). *Cryosphere* **13**, 2259–2280 (2019).

6. 6.

Hanssen-Bauer, I. et al. *Climate in Svalbard 2100 – A Knowledge Base for Climate Adaptation* NCCS Report no 1/2019 commissioned by the Norwegian Environment Agency (Miljødirektoratet) (Norwegian Centre for Climate Services, 2019).

7. 7.

Luncke, B. Luftkartlegningen på Svalbard 1936. *Norsk Geogr. Tidsskr.* **6**, 145–154 (1936).

8. 8.

Blois, J. L., Williams, J. W., Fitzpatrick, M. C., Jackson, S. T. & Ferrier, S. Space can substitute for time in predicting climate-change effects on biodiversity. *Proc. Natl Acad. Sci. USA* **110**, 9374–9379 (2013).

9. 9.

Serreze, M. C. & Barry, R. G. Processes and impacts of Arctic amplification: a research synthesis. *Global Planet. Change* **77**, 85–96 (2011).

10. 10.

Isaksen, K. et al. Recent warming on Spitsbergen – influence of atmospheric circulation and sea ice cover. *J. Geophys. Res. Atmos.* **121**, 11–913 (2016).

11. 11.

Nordli, Ø. et al. Revisiting the extended Svalbard Airport monthly temperature series, and the compiled corresponding daily series 1898–2018. *Polar Res.* (2020).

12. 12.

Nuth, C., Moholdt, G., Kohler, J., Hagen, J. O. & Kääb, A. Svalbard glacier elevation changes and contribution to sea level rise. *J. Geophys. Res. Earth Surf.* **115** (2010).

13. 13.

Dowdeswell, J. A. et al. The mass balance of circum-Arctic glaciers and recent climate change. *Quat. Res.* **48**, 1–14 (1997).

14. 14.

Hagen, J. O., Kohler, J., Melvold, K. & Winther, J.-G. Glaciers in Svalbard: mass balance, runoff and freshwater flux. *Polar Res.* **22**, 145–159 (2003).

15. 15.

Hagen, J. O., Melvold, K., Pinglot, F. & Dowdeswell, J. A. On the net mass balance of the glaciers and ice caps in Svalbard, Norwegian Arctic. *Arct. Antarct. Alp. Res.* **35**, 264–270 (2003).

16. 16.

Möller, M. & Kohler, J. Differing climatic mass balance evolution across Svalbard glacier regions over 1900–2010. *Front. Earth Sci.* **6**, 128 (2018).

17. 17.

Østby, T. I. et al. Diagnosing the decline in climatic mass balance of glaciers in Svalbard over 1957–2014. *Cryosphere* **11**, 191–215 (2017).

18. 18.

Aas, K. S. et al. The climatic mass balance of Svalbard glaciers: a 10-year simulation with a coupled atmosphere-glacier mass balance model. *Cryosphere* **10**, 1089–1104 (2016).

19. 19.

Schuler, T. V. et al. Reconciling Svalbard glacier mass balance. *Front. Earth Sci.* **8**, 156 (2020).

20. 20.

Bjørk, A. A. et al. An aerial view of 80 years of climate-related glacier fluctuations in southeast Greenland. *Nat. Geosci.* **5**, 427–432 (2012).

21. 21.

Kjeldsen, K. K. et al. Spatial and temporal distribution of mass loss from the Greenland Ice Sheet since AD 1900. *Nature* **528**, 396–400 (2015).

22. 22.

Child, S. F., Stearns, L. A., Girod, L. & Brecher, H. H. Structure-from-motion photogrammetry of Antarctic historical aerial photographs in conjunction with ground control derived from satellite data. *Remote Sens.* **13**, 21 (2021).

23. 23.

Dyurgerov, M. B. & Meier, M. F. Twentieth century climate change: evidence from small glaciers. *Proc. Natl Acad. Sci. USA* **97**, 1406–1411 (2000).

24. 24.

WGMS. *Fluctuations of Glaciers Database* (World Glacier Monitoring Service, accessed 15 February 2020).

25. 25.

Beaulieu, C. & Killick, R. Distinguishing trends and shifts from memory in climate data. *J. Clim.* **31**, 9519–9543 (2018).

26. 26.

Sasgen, I. et al. Return to rapid ice loss in Greenland and record loss in 2019 detected by the GRACE-FO satellites. *Commun. Earth Environ.* **1**, 8 (2020).

27. 27.

Velicogna, I., Sutterley, T. & Van Den Broeke, M. Regional acceleration in ice mass loss from Greenland and Antarctica using GRACE time-variable gravity data. *Geophys. Res. Lett.* **41**, 8130–8137 (2014).

28. 28.

Geyman, E. C., Maloof, A. C. & Dyer, B. How is sea level change encoded in carbonate stratigraphy? *Earth Planet. Sci. Lett.* **560**, 116790 (2021).

29. 29.

Sullivan, M. J. et al. Long-term thermal sensitivity of Earth's tropical forests. *Science* **368**, 869–874 (2020).

30. 30.

Liu, J., Wennberg, P. O., Parazoo, N. C., Yin, Y. & Frankenberg, C. Observational constraints on the response of high-latitude northern forests to warming. *AGU Adv.* e2020AV000228 (2020).

31. 31.

Luckman, A. et al. Calving rates at tidewater glaciers vary strongly with ocean temperature. *Nat. Commun.* **6**, 8566 (2015).

32. 32.

Koenderink, J. J. & Van Doorn, A. J. Affine structure from motion. *J. Opt. Soc. Am. A* **8**, 377–385 (1991).

33. 33.

Terrengmodell Svalbard (S0 Terrengmodell) (Norwegian Polar Institute, 2014).

34. 34.

Hay, C. C., Morrow, E., Kopp, R. E. & Mitrovica, J. X. Probabilistic reanalysis of twentieth-century sea-level rise. *Nature* **517**, 481–484 (2015).

35. 35.

Oerlemans, J. *Glaciers and Climate Change* (CRC Press, 2001).

36. 36.

Oerlemans, J. & Fortuin, J. Sensitivity of glaciers and small ice caps to greenhouse warming. *Science* **258**, 115–117 (1992).

37. 37.

Raper, S. C. & Braithwaite, R. J. Low sea level rise projections from mountain glaciers and ice caps under global warming. *Nature* **439**, 311–313 (2006).

38. 38.

Nöel, B. et al. Low elevation of Svalbard glaciers drives high mass loss variability. *Nat. Commun.* **11**, 4597 (2020).

39. 39.

Porter, C. et al. ArcticDEM (2018).

40. 40.

Girod, L., Nielsen, N. I., Couderette, F., Nuth, C. & Kääb, A. Precise DEM extraction from Svalbard using 1936 high oblique imagery. *Geosci. Instrum. Methods Data Syst.* **7**, 277–288 (2018).

41. 41.

Midgley, N. & Tonkin, T. Reconstruction of former glacier surface topography from archive oblique aerial images. *Geomorphology* **282**, 18–26 (2017).

42. 42.

Mertes, J. R., Gulley, J. D., Benn, D. I., Thompson, S. S. & Nicholson, L. I. Using structure-from-motion to create glacier DEMs and orthoimagery from historical terrestrial and oblique aerial imagery. *Earth Surf. Processes Landforms* **42**, 2350–2364 (2017).

43. 43.

Kavan, J. Early twentieth century evolution of Ferdinand glacier, Svalbard, based on historic photographs and structure-from-motion technique. *Geogr. Ann. A* **102**, 57–67 (2020).

44. 44.

Holmlund, E. S. Aldegondabreen glacier change since 1910 from structure-from-motion photogrammetry of archived terrestrial and aerial photographs: utility of a historic archive to obtain century-scale Svalbard glacier mass losses. *J. Glaciol.* **67**, 107–116 (2021).

45. 45.

Loop, C. & Zhang, Z. Computing rectifying homographies for stereo vision. In Proc. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149) Vol. 1, 125–131 (IEEE, 1999).

46. 46.

Bretscher, O. *Linear Algebra with Applications* 5th edn (Pearson, 2013).

47. 47.

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J. & Reynolds, J. M. ‘Structure-from-Motion’ photogrammetry: a low-cost, effective tool for geoscience applications. *Geomorphology* **179**, 300–314 (2012).

48. 48.

McNabb, R., Nuth, C., Kääb, A. & Girod, L. Sensitivity of glacier volume change estimation to DEM void interpolation. *Cryosphere* **13**, 895–910 (2019).

49. 49.

Williams, C. & Rasmussen, C. E. *Gaussian Processes for Machine Learning* Vol. 2, 302, 303 (MIT Press, 2006).

50. 50.

Nuth, C. & Kääb, A. Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change. *Cryosphere* **5**, 271–290 (2011).

51. 51.

Kääb, A. *Remote Sensing of Mountain Glaciers and Permafrost Creep* 48 (Geographisches Institut der Univ., 2005).

52. 52.

James, M. R., Robson, S., d’Oleire Oltmanns, S. & Niethammer, U. Optimising UAV topographic surveys processed with structure-from-motion: ground control quality, quantity and bundle adjustment. *Geomorphology* **280**, 51–66 (2017).

53. 53. Rolstad, C., Haug, T. & Denby, B. Spatially integrated geodetic glacier mass balance and its uncertainty based on geostatistical analysis: application to the western Svartisen ice cap, Norway. *J. Glaciol.* **55**, 666–680 (2009).
54. 54. Granshaw, F. D. & Fountain, A. G. Glacier change (1958–1998) in the north Cascades national park complex, Washington, USA. *J. Glaciol.* **52**, 251–256 (2006).
55. 55. König, M., Nuth, C., Kohler, J., Moholdt, G. & Pettersen, R. *A Digital Glacier Database for Svalbard* 229–239 (Springer, 2014).
56. 56. Nuth, C. et al. Decadal changes from a multi-temporal glacier inventory of Svalbard. *Cryosphere* **7**, 1603–1621 (2013).
57. 57. Huss, M. Density assumptions for converting geodetic glacier volume change to mass change. *Cryosphere* **7**, 877–887 (2013).
58. 58. van Pelt, W. J. J. et al. Multidecadal climate and seasonal snow conditions in Svalbard. *J. Geophys. Res. Earth Surf.* **121**, 2100–2117 (2016).
59. 59. Fürst, J. J. et al. The ice-free topography of Svalbard. *Geophys. Res. Lett.* **45**, 760–769 (2018).

60. 60.

Fürst, J. J. et al. Application of a two-step approach for mapping ice thickness to various glacier types on Svalbard. *Cryosphere* **11**, 2003–2032 (2017).

61. 61.

Morlighem, M. et al. A mass conservation approach for mapping glacier ice thickness. *Geophys. Res. Lett.* **38** (2011).

62. 62.

Hagedoorn, J. & Wolf, D. Pleistocene and recent deglaciation in Svalbard: implications for tide-gauge, GPS and VLBI measurements. *J. Geodyn.* **35**, 415–423 (2003).

63. 63.

Sato, T. et al. A geophysical interpretation of the secular displacement and gravity rates observed at Ny-Ålesund, Svalbard in the Arctic—effects of post-glacial rebound and present-day ice melting. *Geophys. J. Int.* **165**, 729–743 (2006).

64. 64.

Kierulf, H. P., Plag, H. P. & Kohler, J. Surface deformation induced by present-day ice melting in Svalbard. *Geophys. J. Int.* **179**, 1–13 (2009).

65. 65.

Davis, J. L. & Mitrovica, J. X. Glacial isostatic adjustment and the anomalous tide gauge record of eastern North America. *Nature* **379**, 331–333 (1996).

66. 66.

Uppala, S. M. et al. The ERA-40 re-analysis. *Q. J. R. Meteorolog. Soc.* **131**, 2961–3012 (2005).

67. 67.

Reistad, M. et al. A high-resolution hindcast of wind and waves for the North Sea, the Norwegian Sea, and the Barents Sea *J. Geophys. Res. Oceans* **116** (2011).

68. 68.

Førland, E. J. et al. Measured and modeled historical precipitation trends for Svalbard. *J. Hydrometeorol.* **21**, 1279–1296 (2020).

69. 69.

Huybers, P. & Curry, W. Links between annual, Milankovitch and continuum temperature variability. *Nature* **441**, 329–332 (2006).

70. 70.

Amdur, T., Stine, A. & Huybers, P. Global surface temperature response to 11-yr solar cycle forcing consistent with general circulation model results. *J. Clim.* **34**, 2893–2903 (2021).

71. 71.

Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).

72. 72.

Huybers, P. Early Pleistocene glacial cycles and the integrated summer insolation forcing. *Science* **313**, 508–511 (2006).

73. 73.

Benn, D. I. et al. Melt-under-cutting and buoyancy-driven calving from tidewater glaciers: new insights from discrete element and continuum model simulations. *J. Glaciol.* **63**, 691–702 (2017).

74. 74.

Cuffey, K. M. & Paterson, W. S. B. *The Physics of Glaciers* (Academic Press, 2010).

75. 75.

Gabbi, J., Carenzo, M., Pellicciotti, F., Bauder, A. & Funk, M. A comparison of empirical and physically based glacier surface melt models for long-term simulations of glacier response. *J. Glaciol.* **60**, 1140–1154 (2014).

76. 76.

Fujita, K. & Ageta, Y. Effect of summer accumulation on glacier mass balance on the Tibetan Plateau revealed by mass-balance model. *J. Glaciol.* **46**, 244–252 (2000).

77. 77.

Naegeli, K. & Huss, M. Sensitivity of mountain glacier mass balance to changes in bare-ice albedo. *Ann. Glaciol.* **58**, 119–129 (2017).

78. 78.

MacAyeal, D. Binge/purge oscillations of the Laurentide ice sheet as a cause of the North Atlantic's Heinrich events. *Paleoceanography* **8**, 775–784 (1993).

79. 79.

Oerlemans, J. *Minimal Glacier Models* (Utrecht Publishing & Archiving Services, 2008).

80. 80.

Oerlemans, J., Jania, J. & Kolondra, L. Application of a minimal glacier model to Hansbreen, Svalbard. *Cryosphere* **5**, 1–11 (2011).

81. 81.

Florentine, C., Harper, J., Fagre, D., Moore, J. & Peitzsch, E. Local topography increasingly influences the mass balance of a retreating cirque glacier. *Cryosphere* **12**, 2109–2122 (2018).

82. 82.

Mott, R. et al. Avalanches and micrometeorology driving mass and energy balance of the lowest perennial ice field of the Alps: a case study. *Cryosphere* **13**, 1247–1265 (2019).

83. 83.

Kingslake, J. et al. Extensive retreat and re-advance of the West Antarctic Ice Sheet during the Holocene. *Nature* **558**, 430–434 (2018).

84. 84.

Etzelmüller, B. et al. Modeling the temperature evolution of Svalbard permafrost during the 20th and 21st century. *Cryosphere* **5**, 67–79 (2011).

85. 85.

Schuler, T. V. et al. Reconciling Svalbard glacier mass balance. *Front. Earth Sci.* **8**, 156 (2020).

86. 86.

Błaszczyk, M., Jania, J. & Hagen, J. O. Tidewater glaciers of Svalbard: recent changes and estimates of calving fluxes. *Polish Polar Res.* **30** (2009).

87. 87.

Huss, M. & Hock, R. A new model for global glacier change and sea-level rise. *Front. Earth Sci.* **3**, 54 (2015).

88. 88.

Radíć, V. et al. Regional and global projections of twenty-first century glacier mass changes in response to climate scenarios from global climate models. *Clim. Dyn.* **42**, 37–58 (2014).

89. 89.

Marzeion, B., Jarosch, A. & Hofer, M. Past and future sea-level change from the surface mass balance of glaciers. *Cryosphere* **6**, 1295–1322 (2012).

90. 90.

Jacob, T., Wahr, J., Pfeffer, W. T. & Swenson, S. Recent contributions of glaciers and ice caps to sea level rise. *Nature* **482**, 514–518 (2012).

91. 91.

Matsuo, K. & Heki, K. Current ice loss in small glacier systems of the Arctic Islands (Iceland, Svalbard, and the Russian High Arctic) from satellite gravimetry. *Terr. Atmospheric Ocean. Sci.* **24**, 657–670 (2013).

92. 92.

Moholdt, G., Nuth, C., Hagen, J. O. & Kohler, J. Recent elevation changes of Svalbard glaciers derived from ICESat laser altimetry. *Remote Sens. Environ.* **114**, 2756–2767 (2010).

93. 93.

Lang, C., Fettweis, X. & Erpicum, M. Stable climate and surface mass balance in Svalbard over 1979–2013 despite the Arctic warming. *Cryosphere* **9**, 83–101 (2015).

94. 94.

Oerlemans, J. et al. Estimating the contribution of Arctic glaciers to sea-level change in the next 100 years. *Ann. Glaciol.* **42**, 230–236 (2005).

95. 95.

Wouters, B., Gardner, A. S. & Moholdt, G. Global glacier mass loss during the GRACE satellite mission (2002–2016). *Front. Earth Sci.* **7**, 96 (2019).

96. 96.

Box, J. E. et al. Global sea-level contribution from Arctic land ice: 1971–2017. *Environ. Res. Lett.* **13**, 125012 (2018).

97. 97.

Ciraci, E., Velicogna, I. & Swenson, S. Continuity of the mass loss of the world’s glaciers and ice caps from the GRACE and GRACE Follow-On missions. *Geophys. Res. Lett.* **47**, e2019GL086926 (2020).

98. 98.

Radíč, V. & Hock, R. Regionally differentiated contribution of mountain glaciers and ice caps to future sea-level rise. *Nat. Geosci.* **4**, 91–94 (2011).

99. 99.

Straneo, F. et al. Impact of fjord dynamics and glacial runoff on the circulation near Helheim Glacier. *Nat. Geosci.* **4**, 322–327 (2011).

100. 100.

Bonan, D. B., Christian, J. E. & Christianson, K. Influence of North Atlantic climate variability on glacier mass balance in Norway, Sweden and Svalbard. *J. Glaciol.* **65**, 580–594 (2019).

101. 101.

Benn, D., Fowler, A. C., Hewitt, I. & Sevestre, H. A general theory of glacier surges. *J. Glaciol.* **65**, 701–716 (2019).

102. 102.

Dunse, T. et al. Glacier-surge mechanisms promoted by a hydro-thermodynamic feedback to summer melt. *Cryosphere* **9**, 197–215 (2015).

103. 103.

Morris, A., Moholdt, G. & Gray, L. Spread of Svalbard glacier mass loss to Barents Sea margins revealed by CryoSat-2. *J. Geophys. Res. Earth Surf.* **125**, e2019JF005357 (2020).

Acknowledgements

We thank F. Simons, C.-Y. Lai, P. Wennberg, P. Moore, B. Dyer, G. Moholdt, R.A. Morris, E. Isaksson, A. Schomacker, C. Nuth, E. Schytt Holmlund and B. Geyman for conversations that improved the manuscript. W.J.J.v.P. acknowledges funding from the Swedish National Space Agency (project 189/18). E.C.G. was supported by a Daniel M. Sachs Class of 1960 Global Scholarship at Princeton University, a Svalbard Science Forum Arctic Field Grant, and the Fannie and John Hertz Foundation.

Author information

Author notes

1. Emily C. Geyman

Present address: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, USA

Affiliations

1. Norwegian Polar Institute, Fram Centre, Tromsø, Norway

Emily C. Geyman, Harald Faste Aas & Jack Kohler

2. Department of Earth Sciences, Uppsala University, Uppsala, Sweden

Ward J. J. van Pelt

3. Department of Geosciences, Princeton University, Princeton, USA

Adam C. Maloof

Contributions

E.C.G. and J.K. designed the study. E.C.G. performed the SfM reconstructions and the analysis of the data. W.J.J.v.P. assembled and downscaled the regional climate model results. A.C.M. contributed to the formulation of the space-for-time substitution. H.F.A. oversaw the digitization of the 1936/1938 image archive. E.C.G. wrote the manuscript, with edits from all authors.

Corresponding author

Correspondence to [Emily C. Geyman](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Jaime Otero 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 Svalbard's estimated contribution to sea level rise: past and future.

The x -axis represents the total ice loss, expressed in terms of mm of global sea level equivalent (SLE). The SLE is normalized to a 100-yr period to facilitate comparison between historical observations of varying length and projections for the 21st century. For the historical models that only simulate climatic mass balance¹⁹, we added a calving flux⁸⁶ of $-6.75 \pm 1.7 \text{ Gt yr}^{-1}$ to estimate the total mass balance and facilitate comparison with the geodetic and gravimetric observations. Since the data are compiled from many sources^{2,3,4,5, 12,13,14,15,16, 17, 85, 87,88,89,90,91,92,93,94,95,96,97, 98, 103}, the error bars represent different quantities. In most cases, the shaded bars represent the reported $\pm 1\sigma$ uncertainty. For Marzeion et al. (2012), the bars represent the range of results from different ensemble runs, and for ‘This study’, the bars represent the range from using the 5th to 95th percentile of future precipitation estimates (Fig. 4). The 21st century projections for this study are broken into 3 groups: (1) those based on the 1936-2010 mass balance observations using mean summer temperature as the explanatory variable for ice loss (\ddagger ; Figs. 3–4), (2) those based on the 2000-2019 satellite era observations⁴ using mean summer temperature ($*$; Extended Data Fig. 10), and (3) those based on the 1936-2010 observations using positive degree days (\ddagger ; Extended Data Fig. 9). All three methods produce similar estimates for 21st century ice loss. The predictions begin to diverge under the most extreme warming scenario (RCP8.5)⁶, with the space-for-time substitution based on the positive degree day (PDD) approach (Extended Data Fig. 9) producing the highest estimates of sea level rise. Note that the total ice volume of Svalbard is estimated⁵⁹ to be $\sim 6,199 \text{ km}^3$ which is equivalent to 15 mm of sea level rise. Thus, the 21st century predictions exceeding 15 mm of SLE^{87,88,89} are not feasible.

Extended Data Fig. 2 Past and future estimates of temperature and precipitation across Svalbard.

(a-c) The 1,888 glaciers in Svalbard span an elevation range³⁸ of $>1,200$ m, a mean annual temperature range of >10 °C, and a >4 -fold change in precipitation (<0.5 to 2.0 m.w.e. yr^{-1}). The elevation map in (a) uses the Norwegian Polar Institute S0 terrain model³³, and the mean annual air temperature (b) and mean annual precipitation (c) estimates use the downscaled NORA10 dataset⁵ (1 km resolution, 1957-2018). (d-i) Glacier-averaged summer temperature (d-f) and solid precipitation estimates (g-i) for 2010-2100 from Arctic CORDEX under the RCP2.6, RCP4.5, and RCP8.5 scenarios⁶. The color bar for the mean summer temperature (T_s) maps in (d-f) is centered on the upper bound of the 1936-2010 T_s (95th percentile = 2.2 °C; Fig. 3). Thus, brown colors indicate that, in the space-for-time substitution in Fig. 4, we are predicting glacier behavior in response to temperatures that rise higher than those in the observational data that calibrate the model.

Extended Data Fig. 3 An overview of the 1936/1938 aerial survey⁷ and the new datasets available from this study.

(a) Locations of the 5,507 high-oblique aerial photographs acquired during the mapping campaigns⁷ of 1936 and 1938. The coordinates are displayed in WGS84 / UTM zone 33N. (b) We divided the images into 17 groups with convergent camera geometries for SfM reconstruction in Agisoft Metashape (Extended Data Fig. 4). We provide the Svalbard-wide (c) 1936 orthophotomosaic, (d) 1936 digital elevation model (DEM), and (e) 1936-2010 elevation change (Δh) at 20 m and 50 m resolution. We also provide the 1936-2010 Δh datasets at 5 m resolution, although, because of file size constraints, these data are divided into 8 files for each of the Svalbard zones depicted in Fig. 2d. The unprocessed point clouds from the 17 regional SfM reconstructions in (c) are available as .laz files. In addition to the raster datasets in (c-e), we provide a .shp file inventorying the 1936 extents of Svalbard glaciers and a spreadsheet recording statistics such as $\Delta h/\Delta t$, $\Delta M/$

Δt , DEM uncertainty, and NORA10 climate fields⁵ (mean summer temperature, precipitation as snow, etc.). See Data availability.

Extended Data Fig. 4 An overview of the image pre-processing and structure from motion (SfM) pipelines.

To improve feature selection during the SfM reconstruction, we enhance the digitized images by increasing contrast through histogram stretching and sharpening features using the Dehaze Tool in Adobe Lightroom. This radiometric enhancement step improves photogrammetric reconstructions over ice and snow, which tend to be lower contrast than the surrounding land. Finally, since scanning does not preserve the internal geometry (images can be rotated, translated, and warped), we locate the four fiducial marks on the edges of each image and apply a projective transformation that maps the images to a standardized internal geometry. Owing to the large number of images in the dataset, we use an automated pipeline for fiducial mark identification. We convolve the edges of the image with an idealized fiducial template to identify target regions. Next, inside the target regions, we convolve the image with a Laplacian of Gaussian filter to locate the fiducial spot. We process the aerial photographs in a standard photogrammetric workflow in Agisoft Metashape 1.6.0. In brief, we first extract up to 40,000 keypoints from each image. Keypoint matching across all the images provides the constraints to solve for the unknown parameters, including the relative camera locations/orientations and the camera distortion parameters. Adding ground control points (GCPs), with specified (x,y,z) positions, enables the absolute georeferencing of the model. Finally, a multi-view stereo (MVS) reconstruction converts the sparse 3D model to a dense 3D point cloud. We perform the MVS reconstruction with a Dense Quality of Medium (meaning that depth maps are generated at 1/4 the image resolution) and Dense Filtering at Moderate to Aggressive.

Extended Data Fig. 5 Example 3D comparisons of Svalbard glaciers in 1936 and 2008-2011.

(a) Austre/Vestre Brøggerbreen, (b) Midtre/Austre Lovénbreen, (c) Grønfjordbreen, (d) Tungebreen, (e) Gløttfjellbreen, (f) Pedåsenkobreen.

Glacier volume decreased by 11% across Svalbard during the interval 1936–2010 (Fig. 2). The 2008–2011 models use the NPI 5 m regional DEMs³³ and associated 50 cm orthophotos (<https://geodata.npolar.no/>).

Extended Data Fig. 6 Data coverage and void-filling with Gaussian process (GP) regression.

(a) Black areas denote regions with 3D photogrammetric (SfM) constraints from the 1936/1938 aerial images (Fig. 1) and white regions denote void areas infilled with the GP regression (Methods). The SfM-generated point clouds have void areas because of occlusion and poor feature matching in low-contrast areas. There is no reconstruction for the eastern portion of Austfonna (Fig. 2), since no photographs of that region were acquired during the 1936/1938 expeditions⁷ (Extended Data Fig. 3a). (b-e) An illustration of the void filling procedure, applied to Oscar II Land in western Svalbard. To fill the holes in the 1936 DEM, we first compute the Δh map, differencing the ~2010 reference DEM to the 1936 reconstruction (b). Next, we train a GP regression to estimate the Δh values in the void areas. The GP regression is trained using x, y, and z (the 2010 elevation) as predictor variables to infer Δh as the response variable, and thereby incorporates both the spatial information in (a) and the elevation-dependence of Δh illustrated in (c). (d) The error of the GP-regression-infilled values is estimated on random subsets of data points (60%) held-out from model training. Finally, subtracting the infilled Δh map in (e) from the 2010 reference DEM yields the 1936 surface (Extended Data Fig. 3d).

Extended Data Fig. 7 Climatic and geometric controls on glacier mass balance.

(a-b) Temperature control on ice loss. The scatter plots in (a-b) show a similar relationship as Main text, Fig. 3f-g (T_s vs. $\Delta h/\Delta t$), except the y-axis in (a-b) also includes the solid precipitation component. Specifically, $F_{out} = P_{solid} - b$, where P_{solid} is extracted from the downscaled NORA10 dataset⁵, and b is $\Delta h/\Delta t$ converted to m.w.e. yr^{-1} using a density⁵⁷ of 850 kg m^{-3} . Since the glacier-specific precipitation estimates are noisy, the plots in (a-b) show considerably more scatter than those in Main text, Fig. 3f-g. The

advantage of examining the data in terms of F_{out} is that it enables us to extract the physical quantity k_1 , which describes the expected increase in ice loss (m.w.e. yr^{-1}) for each 1°C rise in mean summer temperature. The gray bands in (a-b) depict the 25th–75th percentile uncertainty envelopes of the T_s vs. F_{out} regressions (all glaciers). Note that the ice loss in marine-terminating glaciers is regulated not only by air temperature driving F_{melt} , but also by fjord temperature, bathymetry, and circulation controlling F_{calving} ^{31,73,99}. We take a first-order approach and fit different k constants to land- vs. marine-terminating glaciers. In both (a-b), the k_1 coefficients are larger in land-terminating glaciers than marine-terminating glaciers. This observation of a weaker sensitivity of F_{out} to air temperature is consistent with the F_{out} in marine-terminating glaciers being driven partly by fjord processes that are decoupled from air temperature. Since the satellite-era observations (b) represent a shorter interval and therefore the $\Delta h/\Delta t$ data are more influenced by annual variability¹⁰⁰ and surge cycles¹⁰¹, we only fit glaciers with $\Delta h/\Delta t$ within the 10th–90th percentile range. Glaciers outside this range are depicted with gray dots. Note that the estimated k_1 coefficients from the 1936–2010 observations (a) and satellite-era datasets⁴ (b) are within uncertainty of each other. (c-d) Glacier slope modulates sensitivity to warming. (a) Simple theoretical glacier models⁷⁹ predict that glaciers with steeper slopes should be less sensitive to temperature rise. The rationale is that, for a lower-slope glacier, a given ELA rise of x meters will transfer a larger fraction of the glacier's area from the accumulation zone to the ablation zone³⁸, causing a more substantial decrease in glacier-averaged $\Delta h/\Delta t$. (b) We test whether there is evidence for a bed-slope control on glacier sensitivity to temperature rise⁷⁹ in our 1936–2010 dataset (Fig. 2) by estimating the relationship between T_s and $\Delta h/\Delta t$ for low, medium, and high slope glaciers. The bed slope is calculated as $\Delta z/\Delta x$ of the bed topography⁵⁹ along each glacier's centerline⁵⁶. The distributions in (b) represent the regression slopes derived from weighted total least squares regressions on repeated random 50% subsamples of the dataset.

Extended Data Fig. 8 Visual and quantitative comparison of the regional T_s vs. $\Delta h/\Delta t$ behavior of Svalbard glaciers.

(a-b) A visual comparison between glaciers in (a) NE Spitsbergen and (b) Edgeøya near the end of the 2020 melt season. Note that, in contrast to the glaciers in (a), the glaciers in (b) have no remaining winter snow, as evinced by the darker, debris-rich ice exposed even at the highest reaches of the glaciers. In other words, 100% of the glacier surface lies within the ablation zone. The Sentinel-2 imagery is from August 01, 2020 and the coordinates are in UTM zone 33N. (c-d) Glacier extents in 1936 and 2010. (e-i) Searching for evidence of threshold behavior in the Svalbard-wide 1936–2010 dataset. (e-f) If a strong tipping point already had been reached, such that, at high T_s , glaciers diverged from the linear behavior in Fig. 3, one might expect a T_s vs. $\Delta h/\Delta t$ relationship like that depicted in (f).

However, (e) does not show strong support for the model in (f). Next, we look for a regional signal in the T_s vs. $\Delta h/\Delta t$ relationship. We divide the Svalbard-wide dataset into the 8 regions shown in (h), each of which has a different average T_s (g) and $\Delta h/\Delta t$. (i) For each region, we study the residual between the Svalbard-wide linear T_s vs. $\Delta h/\Delta t$ (e) and the regional observations. The residuals are computed as predicted minus observed, so negative values indicate that the observed mass balance is more negative than the predictions. The regions in (i) are ordered according to the region-averaged glacier bed slope (Extended Data Fig. 7c), from smallest to largest. The glaciers like those in (b) that are committed to a path of pure melting (no accumulation) appear to follow similar T_s vs. $\Delta h/\Delta t$ relationships as the healthier (close to balance) glaciers in (a).

[Extended Data Fig. 9 Glacier sensitivity to warming and 21st century predictions based on positive degree day \(PDD\) estimates.](#)

(a-g) An analogous figure to Fig. 3, except using positive degree days (PDDs) to model the ice loss flux (eqns. 12–13) rather than mean summer temperature. In (g), only bins representing $n \geq 20$ glaciers are shown. (h-k) PDDs are a non-linear function of mean summer temperature. For each glacier, we use the 1957-present daily temperature time series⁵ (h) to understand the relationship between mean summer temperature and PDDs (k). Specifically, we iteratively shift the time series in (h) up at 0.5 °C steps

and compute the new number of PDDs. The red dashed line in (k) depicts what the PDD estimate would be if the melt season didn't get any longer, which is analogous to what the T_s model does (Figs. 3–4). As illustrated in (j), the number of PDDs increases not only because the mean summer temperature rises, but also because the duration of the melt season increases. The plots in (h–k) are for the glacier Bungebreen (76.814°N , 16.097°E). We repeat the analysis for all glaciers on Svalbard to produce glacier-specific relationships between mean summer temperature and PDDs (k). (l) We test the PDD-based space-for-time substitution using 1936–1990 data to calibrate the model, and then compare predictions for the 1990–2010 interval to independent DEM-derived mass balance estimates. We find excellent agreement between the model and observations. Next, we use the model to predict 21st century (2010–2100) mass balance. (m–n) The predictions using the PDD method are nearly identical to the predictions using mean summer temperature (Fig. 4) under the RCP2.6 (m) and RCP4.5 (n) scenarios. However, the PDD method produces more negative mass balance estimates under RCP8.5, because of the divergence of the PDD curve from a simple linear function at higher summer temperatures (k). The brackets in (m–o) show results from simulations using the 5th and 95th percentiles of predicted winter precipitation⁶.

Extended Data Fig. 10 Temperature control on glacier mass balance in the satellite era4 (2000–2019).

An analogous figure to Figs. 3–4, but using 2000–2019 $\Delta h/\Delta t$ derived from ASTER DEMs⁴. The mean summer temperature values are extracted from the downscaled NORA10 dataset⁵. (c) The very negative $\Delta h/\Delta t$ at Basin-3 on Austfonna represents an ongoing surge there with an estimated calving flux¹⁰² of $4.2 \pm 1.6 \text{ Gt yr}^{-1}$, which represents about a quarter of the Svalbard-wide ice loss¹⁰³. (e–g) To reduce the influence of outliers in the regression analysis, we only include the glaciers that have mean summer temperature and $\Delta h/\Delta t$ values that fall within the 10th to 90th percentiles of the Svalbard-wide dataset. (e) As in Fig. 3e, the distribution of regression slopes from the bootstrap resampling scheme does not overlap $0 \text{ m yr}^{-1} \text{ }^\circ\text{C}^{-1}$, indicating a significant temperature dependence of $\Delta h/\Delta t$. The estimated regression slope is $-0.28 [-0.36, -0.22] \text{ m yr}^{-1} \text{ }^\circ\text{C}^{-1}$ (median

with 25-75th percentile range), which is slightly less negative, but still within the uncertainty envelope of the value estimated in Fig. 3 (-0.37 [-0.43 , -0.29] $\text{m yr}^{-1} \text{ }^{\circ}\text{C}^{-1}$). (h) To test whether the 2000-2019 observations⁴ are sufficiently long to characterize the temperature and precipitation control on ice loss (Extended Data Fig. 7), we train a space-for-time substitution using the 2000-2019 observations and the NORA10 temperature and precipitation estimates⁵ for 1936-1990 to estimate mass balance during the period 1936-1990. Comparison with the independent, DEM-derived observations of $\Delta h/\Delta t$ indicates that the space-for-time estimates have an error of 0.05 m yr^{-1} . The simulated 1936-1990 mass balance in (h) may have a larger spread than the observed mass balance due to the somewhat more noisy (affected by surges, interannual variability¹⁰⁰, etc.) satellite-era data used to calibrate the model. (i-k) We use a space-for-time substitution, trained using the 2000-2019 observations⁴ to predict 21st century Svalbard-wide $\Delta h/\Delta t$ under the same three climate scenarios⁶ as in Fig. 4b-d. The brackets show the model runs with the 5th and 95th percentiles of modeled winter precipitation⁶.

Extended Data Table 1 Svalbard glacier changes, 1936-2010

Supplementary information

Supplementary Information

Additional clarification and details about the methodology. The code and the associated datasets required to reproduce the space-for-time analysis can be downloaded from <https://doi.org/10.5281/zenodo.5643856>. The full dataset of 1936/1938 3D glacier reconstructions can be downloaded from <https://doi.org/10.5281/zenodo.5644415>.

Peer Review Information

Supplementary Data

Glacier-by-glacier statistics such as ice loss, climate parameters and twenty-first century predictions.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Geyman, E.C., J. J. van Pelt, W., Maloof, A.C. *et al.* Historical glacier change on Svalbard predicts doubling of mass loss by 2100. *Nature* **601**, 374–379 (2022). <https://doi.org/10.1038/s41586-021-04314-4>

- Received: 17 June 2021
- Accepted: 07 December 2021
- Published: 19 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04314-4>

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

[Future ice loss captured by historical snapshots](#)

- Twila A. Moon

News & Views 19 Jan 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04314-4>

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

- Article
- Open Access
- [Published: 19 January 2022](#)

Global fine-scale changes in ambient NO₂ during COVID-19 lockdowns

- [Matthew J. Cooper](#) ORCID: [orcid.org/0000-0002-4145-3458^{1,2}](https://orcid.org/0000-0002-4145-3458),
- [Randall V. Martin](#) ORCID: [orcid.org/0000-0003-2632-8402^{1,2,3}](https://orcid.org/0000-0003-2632-8402),
- [Melanie S. Hammer](#) ORCID: [orcid.org/0000-0001-8443-7979^{1,2}](https://orcid.org/0000-0001-8443-7979),
- [Pieterernel F. Levelt](#)^{4,5,6},
- [Pepijn Veefkind](#)^{4,7},
- [Lok N. Lamsal](#)^{8,9},
- [Nickolay A. Krotkov](#) ORCID: [orcid.org/0000-0001-6170-6750⁸](https://orcid.org/0000-0001-6170-6750),
- [Jeffrey R. Brook](#)^{10,11} &
- [Chris A. McLinden](#) ORCID: [orcid.org/0000-0001-5054-1380¹²](https://orcid.org/0000-0001-5054-1380)

Nature volume 601, pages 380–387 (2022)

- 2850 Accesses
- 132 Altmetric
- [Metrics details](#)

Subjects

- [Atmospheric chemistry](#)
- [Environmental monitoring](#)

Abstract

Nitrogen dioxide (NO₂) is an important contributor to air pollution and can adversely affect human health^{1,2,3,4,5,6,7,8,9}. A decrease in NO₂ concentrations has been reported as a result of lockdown measures to reduce the spread of COVID-19^{10,11,12,13,14,15,16,17,18,19,20}. Questions remain, however, regarding the relationship of

satellite-derived atmospheric column NO₂ data with health-relevant ambient ground-level concentrations, and the representativeness of limited ground-based monitoring data for global assessment. Here we derive spatially resolved, global ground-level NO₂ concentrations from NO₂ column densities observed by the TROPOMI satellite instrument at sufficiently fine resolution (approximately one kilometre) to allow assessment of individual cities during COVID-19 lockdowns in 2020 compared to 2019. We apply these estimates to quantify NO₂ changes in more than 200 cities, including 65 cities without available ground monitoring, largely in lower-income regions. Mean country-level population-weighted NO₂ concentrations are 29% ± 3% lower in countries with strict lockdown conditions than in those without. Relative to long-term trends, NO₂ decreases during COVID-19 lockdowns exceed recent Ozone Monitoring Instrument (OMI)-derived year-to-year decreases from emission controls, comparable to 15 ± 4 years of reductions globally. Our case studies indicate that the sensitivity of NO₂ to lockdowns varies by country and emissions sector, demonstrating the critical need for spatially resolved observational information provided by these satellite-derived surface concentration estimates.

[Download PDF](#)

Main

Nitrogen dioxide (NO₂) is an important contributor to air pollution as a primary pollutant and as a precursor to ozone and fine particulate matter production. Human exposure to elevated NO₂ concentrations is associated with a range of adverse outcomes such as respiratory infections^{2,3,4}, increases in asthma incidence^{5,6}, lung cancer⁷ and overall mortality^{8,9}. NO₂ observations indicate air quality relationships with combustion sources of pollution such as transportation^{6,21}. Initial investigations found substantial decreases in the atmospheric NO₂ column from satellite observations^{10,11,12,13,14,15,16} and in ambient NO₂ concentrations from ground-based monitoring^{17,18,19,20} during lockdowns enacted to reduce the spread of COVID-19. However, questions remain about the relationship of atmospheric columns with health- and policy-relevant ambient ground-level concentrations, and about the representativeness of sparse ground-based monitoring for broad assessment. Thus, there is need to relate satellite observations of NO₂ columns to ground-level concentrations. It is also important to consider the effect of meteorology on recent NO₂ changes²² and to quantify NO₂ changes due to COVID-19 interventions in the context of longer-term trends²³. Furthermore, air quality monitoring sites tend to be preferentially located in higher-income regions, raising questions about how NO₂ changed in lower-income regions where larger numbers of potentially susceptible

people reside. Estimates of changes in ground-level NO₂ concentrations derived from satellite remote sensing would fill gaps between ground-based monitors, offer valuable information in regions with sparse monitoring, and more clearly connect satellite observations with ground-level ambient air quality.

Previous satellite-derived estimates of ground-level NO₂ used information on the vertical distribution of NO₂ from a chemical transport model to relate satellite NO₂ column densities to ground-level concentrations^{24,25,26}. Recent work improved upon this technique by allowing the satellite column densities to constrain the vertical profile shape, allowing for more accurate representation of sub-model-grid variability, reducing the sensitivity to model resolution and simulated profile shape errors, and improving agreement between the satellite-derived ground-level concentrations and in situ monitoring data²⁷. Applying this technique to examine changes in NO₂ during lockdowns bridges the gap between previous studies focusing on either ground monitors or satellite column densities, thus providing a more complete and reliable picture of the changes in exposure.

Since 2005, the gold standard for satellite NO₂ observations has been the Ozone Monitoring Instrument (OMI) on board NASA's Earth Observing System Aura satellite^{28,29}. The newest remote sensing spectrometer, the European Space Agency's TROPOspheric Monitoring Instrument (TROPOMI)³⁰ on the Copernicus Sentinel 5p satellite, has been providing NO₂ observations with finer spatial resolution and higher instrument sensitivity since 2018. These attributes allow the generation of TROPOMI NO₂ maps at 100 times finer resolution (approximately $1 \times 1 \text{ km}^2$) with a one-month averaging period^{31,32}, an improvement over the spatial and temporal averaging needed for accurate OMI maps (typically approximately $10 \times 10 \text{ km}^2$ over one year)²⁴. Concurrently, the excellent stability of the OMI instrument over the last 15 years provides an ideal dataset for long-term trend analysis^{28,33} that offers context for recent TROPOMI data.

Lockdown restrictions act as an experiment about the efficacy of activity reductions on mitigating air pollution. The Oxford COVID-19 Government Response Tracker (OxCGRT, <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker#data>) has been monitoring government-imposed restrictions, and studies have indicated that NO₂ decreases were larger for cities in countries with strict lockdowns³⁴. However, there is limited information on lockdown stringency on sub-national levels or on how various emission sectors respond to lockdowns. An observation-based metric for lockdown intensity could provide useful information for examining lockdowns on city-level scales or for examining the effects on air quality that are associated with lockdowns in different emission sectors.

Here we leverage the high spatial resolution of TROPOMI to infer global ground-level NO₂ estimates at, to our knowledge, an unprecedented spatial resolution sufficient to assess individual cities worldwide, and to examine changes in ground-level NO₂ occurring during COVID-19 lockdowns from January–June 2020. Case studies presented here demonstrate how the satellite-based estimates provide information on important spatial variability in lockdown-driven NO₂ changes, and in the NO₂ response to lockdowns in various emissions sectors. We also use TROPOMI to provide fine-scale structure to the long-term record of OMI observations (2005–2019), which provides an opportunity to examine trends in ground-level NO₂ over the last 15 years to provide context for the recent changes.

Global NO₂ concentrations and trends

Global annual mean TROPOMI-derived ground-level NO₂ concentrations for 2019 provide an initial baseline (Fig. 1). The excellent resolution ($\sim 1 \times 1 \text{ km}^2$) of ground-level NO₂ concentrations reveal pronounced heterogeneity (Supplementary Figs. 1–7). NO₂ enhancements are apparent over urban and industrial regions. TROPOMI-derived ground-level concentrations exhibit consistency with in situ observations ($r = 0.71$, $N = 3,977$, in situ versus satellite slope = 0.97 ± 0.02), as shown in Supplementary Fig. 8. Neglecting the spatial and temporal variability in the NO₂ column-to-surface relationship degrades the consistency with ground monitors (slope = 0.78 ± 0.01), demonstrating the importance of relating satellite columns to surface concentrations for exposure assessment.

Fig. 1: Satellite-derived ground-level NO₂ concentrations.

 **figure 1**

a, TROPOMI-derived 2019 annual mean ground-level NO₂ concentrations at approximately 1 × 1 km² resolution. **b**, Trend in OMI and TROPOMI-derived annual mean ground-level concentrations from 2005–2019. The colour intensity represents the statistical significance of the trend. **c–e**, Population-weighted mean NO₂ from ground monitors and from satellite-derived NO₂ sampled at ground-monitor locations in China (**c**), Europe (**d**) and North America (**e**), normalized by the mean concentration during the period where ground-monitor data are available. The black (ground-derived) and red (satellite-derived) values give the trends for the period where ground-monitor data are available. Only monitors with data available over the entire time period are included. Error bars represent population-weighted standard deviations. **f**, Population-weighted mean satellite-inferred ground-level NO₂ concentrations in South America, Africa and the Middle East, and Oceania. Trends during the given time periods are given at top. Time periods were chosen to reflect the most recent years where a consistent trend is observed. Error bars represent uncertainties in population-weighted means using a bootstrapping method.

Examination of long-term changes in air pollution offers context for changes during COVID-19 lockdowns (Fig 1, Supplementary Figs. 1–7). Satellite-derived NO₂ concentrations decreased from 2005–2019 in urban areas across most of the USA and Europe, eastern China, Japan, and near Johannesburg, South Africa, largely reflecting emission controls on vehicles and power generation. NO₂ increases are observed in Mexico, the Alberta oil sands region in northern Canada, and throughout the Balkan peninsula, central and northern China, India and the Middle East, broadly consistent with reported trends in ground-monitor data^{35,36,37}. Trends in China can be separated into three regimes: ground-level concentrations increased in China from 2005–2010, plateaued from 2010–2013, and decreased from 2013–2019. This change was driven by stricter vehicle and power generation emission standards³⁸ and is consistent with observed changes in NO₂ columns^{39,40}. Similarly, concentrations increased in urban and industrial areas of South America from 2005–2010, and in South Africa and the Middle East from 2005–2015, and decreased in more recent years. Maps of trends in these regions for these time periods are shown in Supplementary Fig. 9. Concentrations in India increased across both time periods owing to increasing coal-powered electricity demands and growing industrial emissions⁴¹. Trends in population-weighted NO₂ concentrations, used to represent population NO₂ exposure, were calculated using ground monitors and coincidentally sampled satellite observations in North America, Europe and China. Satellite-derived concentrations exhibit decreasing trends ($-2.8 \pm 0.2\% \text{ yr}^{-1}$ in Europe 2005–2019, $-4.3 \pm 0.7\% \text{ yr}^{-1}$ in North America 2005–2019, and $-6.0 \pm 0.7\% \text{ yr}^{-1}$ in China 2015–2019) that agree well with trends in the ground-monitor data (within $0.7\% \text{ yr}^{-1}$ in North America, $0.3\% \text{ yr}^{-1}$ in Europe, and $1.2\% \text{ yr}^{-1}$ in China).

Regional NO₂ changes during lockdowns

Figure 2 shows the April 2020 to April 2019 difference between mean ground-level NO₂ concentrations derived from TROPOMI observations. NO₂ concentrations are lower in most regions in 2020 than in 2019, particularly over urban areas, with global population-weighted mean concentrations decreasing by 16% in 2020 relative to 2019. Fig. 3 shows regional maps focusing on the month with the largest change in population-weighted regional mean concentration for each region, with an additional period included for China, as lockdown restrictions occurred earlier than in other countries. Regional population-weighted mean concentrations decreased by 17–43%. The largest decreases occur in China in February with concentration decreases exceeding 10 parts per billion by volume (ppbv) and substantial decreases persisting in eastern urban areas through April. Thus these lockdown measures temporarily bolstered the decreasing trends across North America⁴² and Europe²⁵ over the last two decades and in China since 2012⁴³, owing to technological advances in vehicles and

power generation, while temporarily buffering changes from increasing energy demands in India and the Middle East^{[40,44,45](#)}. NO₂ increases in April 2020 in central China (Chengdu and Chongqing) because lockdowns began lifting during this time.

Fig. 2: Differences in April mean ground-level NO₂ from 2020 to 2019.

 figure 2

Concentrations derived using TROPOMI observations gridded at approximately 1 × 1-km² resolution.

Fig. 3: Changes in ground-level NO₂ during lockdowns.

 **figure 3**

Left in each pair of images, TROPOMI-derived monthly mean NO₂ differences from 2020–2019 at approximately 1 × 1 km². Right, OMI+TROPOMI-derived NO₂ trends. Annual mean long-term trends are corrected for seasonal variation. The time periods for trend calculations in each region were chosen to reflect the most recent years where a consistent trend is observed and are indicated above the maps. Value under each panel represents population-weighted mean difference for the given region.

Figure 3 shows maps of long-term NO₂ trends for context. In most regions, the observed changes during COVID-19 restrictions exceed the expected year-to-year differences observed in the long-term trends (Table 1). 2020–2019 population-weighted mean concentration changes are lower than long-term trends by factors of 17 ± 7 in North America, 19 ± 2 Europe, of 2.9 ± 0.6 in Africa and the Middle East, of 3.6 ± 0.6 in Asia, 8 ± 7 in South America, and 2 ± 2 in Oceania.

Table 1 TROPOMI-derived, population-weighted ground-level NO₂ data

Meteorological differences are calculated with the GEOS-Chem chemical transport model using emission inventories that do not include changes that occurred owing to COVID-19 lockdown policies but do reflect meteorological changes. Supplementary Fig. 10 shows TROPOMI-derived changes at $2.0^\circ \times 2.5^\circ$ resolution for comparisons with simulated values at the same resolution. Population-weighted NO₂ concentration changes due to meteorology in Asia, Europe, South America, Africa and the Middle East are a factor of 2–6 smaller than observed; thus, meteorology alone cannot explain the observed decreases. Concentration increases in the central USA, as noted in other studies¹⁰, do not appear to be meteorologically driven and may be due to changes in biogenic NO_x sources.

Supplementary Fig. 11 shows the ratio of population-weighted January–June monthly mean NO₂ concentrations in 2020 to 2019 across selected regions. Most regions have the largest decrease in NO₂ in April when lockdown conditions were strongest (the global mean COVID restriction stringency index (defined in Methods) reached a maximum of 0.79 on 18 April), apart from China, where lockdowns were initiated in January. In most regions, 2020 NO₂ concentrations return towards pre-lockdown values in May or June owing to relaxing travel restrictions (30 June global mean stringency index, 0.60) as well as increasing soil, lightning and biomass-burning emissions that lessen the sensitivity of ambient NO₂ to anthropogenic emissions.

City- and country-level NO₂ changes

The fine resolution of our satellite-derived ground-level NO₂ dataset enables the assessment of larger changes in NO₂ concentrations from 2020–2019 evident at the city level. We calculate changes in TROPOMI-observed monthly mean ground-level NO₂ from 2020–2019 over 215 major cities (the ten most populous cities in each country with a population greater than 1 million) for the month with the greatest monthly mean lockdown stringency index, compared with expected changes due to meteorology and long-term trends (Supplementary Table 1). Most cities have TROPOMI-derived NO₂ decreases that cannot be explained by changes due to

meteorology alone. For example, satellite-derived NO₂ concentrations in Beijing decreased by 45% in March, despite meteorological conditions favourable to increased NO₂. Jakarta, Manila, Istanbul, Los Angeles and Buenos Aires among others had decreased NO₂, despite similarly unfavourable meteorological conditions. Some cities, including Moscow, Tokyo, London, New York, Toronto and Delhi, had meteorological conditions that would have led to NO₂ decreases regardless of emission changes, but observed concentration changes exceeded the expected meteorological change.

Consistent analysis of individual cities as enabled by this dataset reveals a mean observed decrease of $32 \pm 2\%$ for these 215 cities. The mean expected meteorologically driven change was $-1 \pm 1\%$ and the mean expected change owing to long-term trends was a decrease of $1.4 \pm 0.4\%$. Supplementary Fig. 12 shows these reductions to be consistent with those found in 381 ground-monitor values from 79 studies³⁴ ($32 \pm 2\%$). Of the 215 cities included here, 65 are in countries that did not have ground-monitoring data available for previous studies. Notably, the 65 cities without monitors are largely in lower-income countries of Africa and southeast Asia. The average gross national income per capita for unmonitored countries is US\$7,100, compared to US\$25,000 for monitored countries, illustrating the potential of satellite-derived ground-level concentrations for providing information about lower-income regions. In summary, the observed decreases in NO₂ across more than 200 cities worldwide were generally driven by COVID-19 lockdowns, with locally varying modulation by meteorology and business-as-usual changes.

Table 1 shows monthly mean country-level population-weighted NO₂ concentrations, changes during COVID-19 lockdown restrictions, meteorological effects and long-term trends for the month with the greatest 2020–2019 change. Meteorological effects were generally minor at the national and regional scale. Multi-year trends provide context for the scale of the changes observed during COVID-19 lockdowns. The decrease in March NO₂ concentrations in the USA from 2019 to 2020 was equivalent to four years of long-term NO₂ reductions. Similarly, changes in NO₂ during COVID-19 lockdowns were equivalent to greater than three years of reductions in China, and up to 23 years in Germany. Globally, the April 2020 population-weighted NO₂ concentration was 0.53 ± 0.06 ppbv lower than in April 2019, equivalent to 15 ± 4 years of global NO₂ reductions.

NO₂ as a lockdown indicator

The relationship between this satellite-derived ground-level NO₂ dataset and lockdown stringency provides supporting evidence for the impact of travel restrictions (Supplementary Fig. 13). The ratio of population-weighted mean observed NO₂ in

2020 to 2019 was calculated for each country and each month from January to June. The 2020/2019 NO₂ ratio in countries with the strictest lockdown (monthly minimum stringency indices greater than the 75th percentile) was $29 \pm 3\%$ lower than for countries with the weakest lockdowns (monthly median stringency indices less than the 25th percentile). Maximum and median ratios were also lower for countries with strict lockdowns. Both distributions have similar variability (standard deviations 0.02 and 0.03) which demonstrates similar interannual variability due to meteorology for both sets. When focusing on only the month with the strictest lockdown for each country, changes in population-weighted NO₂ are correlated with lockdown intensity, with changes in countries with strict lockdowns (average decrease 43% if lockdown index >80) more than three times as large as in those with weaker lockdowns (12% if lockdown index <40).

This relationship suggests that changes in satellite-derived NO₂ concentrations offer observational information on the spatial distribution of lockdown effects that is not available through policy-based stringency indices. For example, although the policy-based stringency index in most cases provides a single value for a country, city-level NO₂ concentration decreases in India are in the range 30–84%, reflecting variability in local mobility restrictions, emissions sources, and their sensitivity to lockdowns. Supplementary Fig. 14 explores the sensitivity of NO₂ concentrations to emissions from the transportation and electricity sectors in India, China and the USA by examining the distribution of changes in NO₂ concentration at the 20 largest population centres and 20 largest fossil fuel-burning power plants in each country. All countries have substantial NO₂ decreases in cities, but the sensitivities vary in areas associated with the electricity sector, with decreasing concentrations near power plants in India (mean change $-35 \pm 4\%$) and China ($-28 \pm 8\%$) but insignificant changes in the USA ($-4 \pm 8\%$). Observed NO₂ changes at these power plants exceed expected changes from meteorology alone ($-8 \pm 2\%$, $-1 \pm 4\%$ and $-1 \pm 3\%$ in India, China and the USA, respectively). Although variability between power plants reflects a mix of regionally varying factors, including meteorology, electricity demand, fuel type and plant-specific emission controls, as well as changes in nearby emissions from other sectors including transportation, these differences indicate a sensitivity of local air quality to activity restrictions affecting the energy sector.

Examining geographic differences in satellite-derived NO₂ concentrations within metropolitan regions is also informative. For example, variability between emission sources is apparent around the city of Atlanta, Georgia, USA (Supplementary Fig. 15). The population-weighted NO₂ concentration in Atlanta and the surrounding region dropped by 28% from April 2019 to 2020, but with substantial spatial variability in the observed change. The greatest NO₂ decreases are found near a large coal-powered electricity plant to the southeast of the city, with significant changes near another plant

to the northwest. Decreases were also larger near the Hartsfield–Jackson International Airport—reflecting the dramatic slowdown in air travel—and over suburban regions to the west and northeast of the city centre, than in the downtown core. Supplementary Fig. 15 also demonstrates the range of NO₂ changes experienced by the local population. Over 1.2 million people live in regions where NO₂ decreases exceeded 40%, whereas nearly 1 million people experienced decreases of 10% or less. Similar heterogeneity in population exposure exists in other major cities, as demonstrated by Supplementary Fig. 16. For example, a subset of over 1 million people in the Paris metropolitan area experienced NO₂ decreases of 75% (4.5 ppbv) or more (10th-percentile exposure), whereas another similar-sized subset experienced changes of 23% (0.6 ppbv) or less (90th-percentile exposure). Of the cities examined here, 68 had an interquartile range in population exposure change during lockdowns of 20 percentage points or larger, 22 of which were unmonitored cities. Studies have found that NO₂ changes during lockdowns varied among socioeconomic, ethnic and racial groups in US cities⁴⁶, and thus the variability in other major cities observed here suggest similar disparities may occur elsewhere. The heterogeneity of NO₂ changes demonstrates the need for the finely resolved information on lockdown effects that is offered by satellite observations.

We find that using this satellite-derived NO₂ dataset as an observational proxy for lockdown conditions is also useful for identifying links between lockdown-driven emission changes and secondary pollutants. For example, several studies have found little to no change in fine particulate matter (PM_{2.5}) during lockdowns as meteorology, long-range transport and nonlinear chemistry complicate the relationship between PM_{2.5} and NO_x emissions^{47,48}. A challenge in these studies has been limited observational information on the local lockdown intensity. Recent work examining 2020–2019 changes in satellite-derived PM_{2.5} concentrations found that lockdown-driven decreases in PM_{2.5} concentration can be identified by separating the meteorological effects from emissions effects using chemical transport modelling and focusing on regions with the greatest sensitivity to emission reductions⁴⁹. Here we examine that same satellite-derived PM_{2.5} dataset using TROPOMI-derived ground-level NO₂ concentrations to identify the regions where PM_{2.5} concentrations are most likely associated with lockdowns or sensitive to NO_x emissions. Supplementary Fig. 17 shows the distribution of changes in monthly mean PM_{2.5} concentrations from 2020–2019 for China in February and North America and Europe in April. Regions with the largest 2020–2019 NO₂ concentration decreases (90th percentile) are considered to be those with significant NO_x emission reductions. Population-weighted mean PM_{2.5} concentrations decreased overall; however, regions with the largest NO₂ decreases experienced greater local changes in PM_{2.5} concentration in China and to a

lesser extent in North America, indicating that the sensitivity of PM_{2.5} to changing NO_x emissions can be inferred. The year-to-year variability of PM_{2.5} concentrations in Europe is similar regardless of changes in NO₂, indicating a greater role of meteorology or transport on PM_{2.5} in this region and period. These results are consistent with previous findings when using chemical transport modelling to identify regions where local emissions are important⁴⁹. Thus, the observational proxy on lockdown conditions offered by these satellite-derived surface NO₂ concentrations offers spatially resolved information to identify where PM_{2.5} and NO₂ (and by proxy, NO_x emissions) are most strongly coupled.

Implications

The pronounced decreases in ground-level NO₂ found here for over 200 cities worldwide during COVID-19 lockdowns are a culmination of recent advancements in techniques for estimating ground-level NO₂ from satellite observations²⁷ alongside higher-resolution satellite observations from TROPOMI that allow for estimating high spatial resolution, short-term changes in NO₂ exposure. This method bridges the gap between monitor data (that measure ground-level air quality but have poor spatial representativeness) and satellite column data (that provide spatial distributions but are less representative of ground-level air quality). The ability to infer global ground-level NO₂ concentrations with sufficient resolution to assess individual cities and even within-city gradients is an important development in satellite remote-sensing instrumentation and algorithms. Additionally, these satellite-derived ground-level NO₂ concentrations offer information about unmonitored communities and populations that are underrepresented in studies focused on ground-monitor data. These cities are found to have different characteristics of NO₂ concentrations and changes during lockdowns that motivate the need for satellite observations in the absence of local ground monitoring. The changes in ground-level NO₂ due to COVID-19 lockdown restrictions, which exceed recent long-term trends and expected meteorologically driven changes, demonstrate the impact that policies that limit emissions can have on NO₂ exposure. This information has relevance to health impact assessment; for example, studies focused on ground-monitor data have indicated improvements in health outcomes related to improved air quality during lockdowns, including an estimated 780,000 fewer deaths and 1.6 million fewer paediatric asthma cases worldwide due to decreased NO₂ exposure²⁰. Our study demonstrates considerable spatial variability in lockdown-driven ground level NO₂ changes that does not necessarily correlate with population density, demonstrating probable uncertainties arising from extrapolating changes observed by ground monitors to estimate broad

changes in population NO₂ exposure. Satellite-based ground-level NO₂ estimates provide high-resolution information on the spatial distribution of NO₂ changes in 2020 that cannot be achieved through ground monitoring, particularly in regions without adequate ground monitoring, and should improve exposure estimates in future health studies. Additionally, ground-level concentrations from downscaled OMI observations provide the opportunity to contrast effects of past mitigation efforts on long-term NO₂ trends against the short-term changes resulting from more dramatic regulations, and a chance to improve studies of health outcomes related to long-term NO₂ exposure.

The strength of the links between observed changes in NO₂ concentration and lockdown stringency indicates that satellite-based ground-level NO₂ concentrations offer useful observational, spatially resolved information about lockdown conditions. This provides an observational metric for examining the efficacy of lockdown restrictions on restricting mobility for studies examining the spread of COVID-19. Here we exploited this information to illustrate the differing sensitivity of NO₂ concentrations to changes in various emission sources to lockdown restrictions. Future applications of these data could include examining socioeconomic drivers that impact this variability within and between countries. Comparisons between satellite-derived ground-level NO₂ and PM_{2.5} also indicate the utility of these data as an observational proxy for identifying regions where secondary pollutants such as PM_{2.5} or ozone are more likely to be sensitive to NO_x emissions; these links are otherwise difficult to trace without the use of chemical transport models⁵⁰.

These data offer information to improve NO₂-exposure estimates, to examine exposure trends, and subsequently estimate changes in health burden. These developments provide an excellent opportunity for advances in air quality health assessment in relation to NO₂ and its combustion-related air pollutant mixture.

Methods

Data

We use tropospheric NO₂ columns from the OMI (NASA Standard Product version 4)⁵¹ and TROPOMI^{52,53} satellite instruments. Both instruments measure solar backscatter radiation in the ultraviolet-visible (UV-vis) spectral bands on sun-synchronous orbits with local overpass times around 1:30 p.m. TROPOMI observations from April 2018–October 2020 are used to examine near-term NO₂, and OMI observations from January 2005–December 2019 are used to examine long-term trends. Observations with retrieved cloud fractions greater than 0.1 or flagged as poor

quality or snow-covered (that is, TROPOMI quality assurance flag <0.75) are excluded. Although the resolution of TROPOMI observations is $3.5 \times 5.5 \text{ km}^2$, several studies have demonstrated that oversampling techniques can provide accurate NO₂ maps at $1 \times 1 \text{ km}^2$ resolution when averaging over a one-month period^{31,32,54}. An area-weighted oversampling technique^{55,56} is used to map daily satellite NO₂ column observations from TROPOMI onto a $\sim 0.01^\circ \times 0.01^\circ$ ($\sim 1 \times 1 \text{ km}^2$) resolution grid and from OMI to a $0.1^\circ \times 0.125^\circ$ ($\sim 10 \times 10 \text{ km}^2$) grid, as these resolutions balance the need of fine resolution for observing fine-scale structure and of minimizing the effects of sampling biases and noise in the observations. Supplementary Fig. 8 provides further evidence that a one-month period provides sufficient observations for a $1 \times 1 \text{ km}^2$ map as the agreement between TROPOMI-derived surface concentrations and in situ observations does not deteriorate when the sampling period is reduced from one year to one month. Additionally, we compared 2019 monthly mean concentration estimates with the 2019 annual mean and find high correlation ($r = 0.90$), indicating similar spatial variability. We correct for sampling biases in the satellite records due to persistent cloudy periods or surface snow cover using a correction factor calculated with the GEOS-Chem chemical transport model described below by sampling the GEOS-Chem-simulated monthly or annual mean column densities to match the satellite.

We use hourly ground-level NO₂ measurements from monitors to constrain and evaluate the satellite-based estimates. Observations from the US Environmental Protection Agency Air Quality System (https://aqs.epa.gov/aqsweb/documents/data_mart_welcome.html) over the continental USA from 2005–2020, Environment and Climate Change Canada’s National Air Pollution Surveillance Program (<http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx>) from 2005–2019, European Environment Agency (<https://aqportal.discomap.eea.europa.eu/index.php/users-corner/>) from 2005–2020, National Air Quality Monitoring Network in China from 2015–2020 were (obtained from <https://quotsoft.net/air>) were used. European monitors classified as near-road are excluded. Monthly and annual mean concentrations at each site are calculated by averaging hourly observations between 13:00–15:00 h (corresponding to satellite overpass times) and corrected for the known overestimate in regulatory measurements due to interference of other reactive nitrogen species following Lamsal et al.²⁴.

To examine the relationship between COVID-19 lockdown policies and ground-level NO₂ concentrations, we use the Oxford COVID-19 Government Response Tracker (OxCGRT, <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker#data>). OxCGRT provides a daily country-level policy ‘stringency index’ ranging from 0–100 that is based on containment and closure policies (for example, school and workplace closures, stay-at-home orders, gathering

restrictions). We also use population density data from the Center for International Earth Science Information Network for the available years of 2005, 2010, 2015 and 2020, and linearly interpolate for other years (<https://doi.org/10.7927/H4JW8BX5>).

Inferring ground-level concentrations from satellite column observations

Ground-level NO₂ concentrations are derived from TROPOMI NO₂ columns following the method developed in Cooper et al.²⁷. This algorithm builds upon the method first developed by Lamsal et al.²⁴ which uses the GEOS-Chem-simulated relationship between ground-level and tropospheric column NO₂ concentrations. The updated algorithm uses the satellite-observed column densities and ground-monitor data as observational constraints on the shape of the boundary layer profile, reducing the sensitivity to model resolution and improving agreement between satellite-derived ground-level concentrations and in situ observations. Technical details on the application of this method as used here are available in the Supplementary Information.

For long-term trend analysis, we use more recent TROPOMI observations to provide fine-resolution spatial structure to the OMI-observed NO₂ columns following the method of Geddes et al.²⁵. Annual mean OMI NO₂ columns are gridded to 10×10 km² resolution and a median-value filter is applied to reduce noise. We smooth the two-year (April 2018–April 2020) mean TROPOMI NO₂ columns mapped at 1×1 km² resolution using a two-dimensional boxcar algorithm with an averaging window of 10×10 km² to match the resolution of the gridded OMI NO₂ columns. We then downscale the annual mean OMI NO₂ columns using the ratio of the 1×1 km² TROPOMI columns to the smoothed TROPOMI columns. The downscaled columns are then used to infer ground-level concentrations following the method used for TROPOMI. Supplementary Fig. 18 demonstrates the utility of this downscaling approach by comparing OMI-derived ground-level concentrations to those derived from the downscaled columns. When comparing 2020–2019 changes in monthly mean concentrations to long-term trends, trends in annual mean concentration are scaled by the ratio of the 2019 monthly mean to the 2019 annual mean to account for seasonality.

The GEOS-Chem chemical transport model version 11-01 is used here (<https://geos-chem.seas.harvard.edu/>) for NO₂ vertical profiles and to assess meteorological effects. GEOS-Chem simulates atmospheric chemistry and physics using a detailed HO_x–NO_x–VOC–O₃–aerosol chemical mechanism^{57,58} driven by meteorological data from the MERRA-2 Reanalysis of the NASA Global Modeling and Assimilation Office⁵⁹.

A detailed description of the simulation is provided in Hammer et al.⁶⁰. We replace the a priori profile used in the retrieval with profiles simulated using the GEOS-Chem model to ensure consistency in vertical profile representation between TROPOMI, OMI, and GEOS-Chem. We simulate NO₂ profiles from January 2005–June 2020 at a horizontal resolution of 2° × 2.5°. Supplementary Fig. 19 shows results from tests using a simulation at 0.5° × 0.625° which was available over North America, Europe and Asia. Satellite-derived ground-level concentrations at ~1 × 1 km² resolution were not sensitive to the resolution of the a priori information, consistent with Cooper et al.²⁷, and thus the 2° × 2.5° was used here for consistency across all regions.

Inferring country- and city-level NO₂ changes during COVID lockdowns

City-level monthly means are calculated from TROPOMI-derived concentrations at ~1 × 1 km² resolution averaged over a 20 × 20 km² region surrounding the city. Meteorological effects are estimated using GEOS-Chem simulations at 2° × 2.5° resolution with consistent emissions in both years, downscaled to ~1 × 1 km² resolution using the horizontal variability of TROPOMI-derived ground-level concentrations. Supplementary Fig. 20 demonstrates that GEOS-Chem simulations can represent meteorologically driven changes in NO₂ in pre-lockdown periods. Trends are defined over 2005–2019 for North America, Europe and Australia, 2015–2019 for Asia and Africa, and 2010–2019 for South America and scaled for seasonality.

Country-level population-weighted means, used to represent population NO₂ exposure, are calculated using concentrations at ~1 × 1-km² resolution via:

$$\text{mean} = \frac{\sum_{i=1}^n P_i x_i}{\sum_{i=1}^n P_i} \quad (2)$$

where x_i is the NO₂ concentration and P_i is the population within a ~1 × 1-km² grid box.

Limitations and sources of uncertainty

Uncertainty values for country- and region-level population-weighted means (σ_{total}) represent the sum in quadrature of three main error sources:

$$\text{Equation 3: } \sigma_{\text{total}} = \sqrt{\sigma_{\text{pop}}^2 - \sigma_{\text{weighted}}^2 + \sigma_{\Omega_{\max}}^2} + \sigma_{\text{AMF2020}}$$

Uncertainty in population-weighted means ($\sigma_{\text{pop-weighted}}$) are estimated using a bootstrapping method⁶¹. Uncertainty in 2020 NO₂ estimates (σ_{AMF2020}) arises from the use of simulated profiles as a priori information for calculating satellite air mass factors and for informing the column-to-ground-level relationship, as these simulations use emission inventories that do not reflect changes resulting from COVID-19-related travel restrictions. Such errors may result in overestimating the fraction of columnar NO₂ near the surface, resulting in an overestimate in satellite-derived ground-level NO₂ concentrations and an underestimate of the 2020–2019 difference. We estimate σ_{AMF2020} by performing sensitivity studies where anthropogenic NO_x emissions were uniformly reduced by 50% to assess the effect of such emission errors on ground-level NO₂ estimates. Reducing anthropogenic NO_x emissions by 50% led to a 5% change in monthly mean population weighted NO₂ concentrations in North America, Europe and Asia for March 2020. Aerosols can also contribute to uncertainty in air mass factor calculations, as a reduction in anthropogenic scattering aerosols during lockdowns may reduce air mass factors leading an underestimation of the NO₂ change^{62,63}. However, this is likely to be a minor source of uncertainty in estimated NO₂ changes due to lockdown, because aerosol concentration changes were small in most regions⁴⁹ and a reduction in aerosol concentration of 10% translates to an uncertainty in NO₂ of less than 5%⁶⁴. Additional uncertainty ($\sigma_{\Omega_{\max}}$) may arise from the choice of the Ω_{\max} parameter (described in the Supplementary Information), particularly in regions where there are insufficient ground-monitor data for constraining Ω_{\max} . We estimate $\sigma_{\Omega_{\max}}$ by evaluating the sensitivity of mean population-weighted NO₂ concentrations to a 20% change in Ω_{\max} . Median country-level $\sigma_{\Omega_{\max}}$ values are ~7%. Uncertainty values in trends are calculated by a weighted linear regression where annual mean concentrations are weighted by σ_{total} .

Although tests here indicate that satellite-derived ground-level NO₂ concentrations are insensitive to the resolution of the simulated data used in the algorithm, discontinuities can occur at the edges of simulation grid boxes. To quantify this uncertainty, we calculate the difference across the grid box boundaries in each region. In most regions the discontinuity is small (<0.5 ppbv in 92% of total cases, and in 98% of cases where NO₂ concentrations >2 ppbv) although can be larger in some cases (>2 ppbv in 0.02% of cases where NO₂ concentrations >2 ppbv, maximum of 4.5 ppbv).

The along-track resolution of TROPOMI observations changed from 7 km to 5.5 km in August 2019. This change may influence interannual comparisons, particularly with respect to the sub-grid downscaling of process which relies on the spatial structure observed by the satellite. To test the influence of this change, we perform a case study where annual mean surface concentrations over Asia are calculated using two different sub-grid scaling factors (v in equation S1 in the Supplementary Information) determined from one year of observations before and after the resolution change, with other variables held constant. The mean relative difference between the two tests was 9% for grid boxes with annual mean concentrations greater than 1 ppbv, with a change in regional population-weighted NO₂ concentrations of 3%. Greater sensitivity to observation resolution was evident in regions with larger NO₂ enhancements, although relative differences greater than 25% occur in fewer than 5% of grid boxes. These tests indicate that although the change in observation resolution may change some spatial gradients, the overall impact on population exposure estimates is small.

Uncertainty values presented above represent uncertainty in the conversion of satellite-observed slant columns into surface concentrations and do not represent systematic errors in the retrieval of slant columns from satellite-observed radiances (~10%), or errors in the air mass factor calculations (23–37%), both of which have been extensively examined in prior studies^{52,65}. Errors related to air mass factor calculations can be reduced by using higher-resolution inputs in air mass factor calculations^{66,67} and are partially mitigated here during the conversion of column densities to surface concentrations through the sub-grid parameterization²⁷.

Although we apply a scaling factor to correct for sampling biases due to persistent cloud cover or surface snow cover, biases in monthly mean calculations may persist if the sampling rate is sufficiently low, particularly for city-level calculations. Most of the cities examined in Supplementary Table 1 had sufficient sampling to allow for a robust monthly mean calculation (median sampling rate of 14 days per month for the months indicated in the table), except for two cities for which fewer than 5 days of observations per month were available for the given month in either 2019 or 2020 (labelled * in Supplementary Table 1). However, results from these cities were consistent with nearby, more frequently sampled cities, lending confidence to these results despite the lower sampling frequency.

This dataset represents substantial improvement over past satellite-derived ground-level NO₂ estimates, as the updated algorithm is less sensitive to model resolution and leverages higher-resolution satellite observations than previous estimates. However, limitations remain. There can be considerable fine-scale variability at scales finer than the $1 \times 1 \text{ km}^2$ resolution used here that cannot be captured by the satellite observations^{68,69}. Additionally, ground-monitor data are used as a constraint in converting observed column densities to ground-level concentrations, and thus

absolute concentration values are probably less accurate in time periods or regions where ground-monitor data are unavailable. However, these data are still useful for examining relative interannual variability or trend analysis. In combining OMI and TROPOMI observations we assume that the spatial gradients observed by TROPOMI in 2018–2020 can be applied to OMI for the entire 2005–2019 time series. New or disappearing point emission sources with small plume footprints may affect this assumption; however, past evaluations of similar assumptions have not found it to be a substantial error source²⁵. Additional errors in the column to ground-level conversion may occur in areas with substantial free tropospheric NO₂ sources such as aircraft emissions or lightning.

Data availability

TROPOMI-derived 2019 annual mean ground-level NO₂ concentrations developed here are available at <https://doi.org/10.5281/zenodo.5484305>. TROPOMI-derived January–June 2019 and 2020 concentrations are available at <https://doi.org/10.5281/zenodo.5484307>. Satellite-derived ground-level NO₂ concentrations for 2005–2019 used for trend analysis are available at <https://doi.org/10.5281/zenodo.5424752>. Satellite column data used here are available from the NASA Goddard Earth Sciences Data and Information Services Center (TROPOMI, <https://doi.org/10.5270/S5P-s4ljjg54>; OMI, 10.5067/Aura/OMI/DATA2017). The GEOS-Chem model version used here is available at <https://doi.org/10.5281/zenodo.2658178>. Hourly ground-level NO₂ measurements from ground monitors in the USA are available from the US Environmental Protection Agency Air Quality System (https://aqs.epa.gov/aqsweb/documents/data_mart_welcome.html), in Canada from Environment and Climate Change Canada’s National Air Pollution Surveillance Program (<http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx>), in Europe from the European Environment Agency (<https://aqportal.discomap.eea.europa.eu/index.php/users-corner/>), and in China from <https://quotsoft.net/air>. COVID-19 lockdown policy information is provided by the Oxford COVID-19 Government Response Tracker (<https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker#data>). Population distribution data are available from the Center for International Earth Science Information Network, <https://doi.org/10.7927/H4JW8BX5>. NO₂ changes during COVID-19 lockdowns from previous studies used for comparison here were compiled by Gkatzelis et al.³⁴ and are available at <https://covid-aqs.fz-juelich.de>. Gross National Income data were provided by World Bank, available at https://data.worldbank.org/indicator/ny.gnp.pcap.cd?year_high_desc=true.

Code availability

Code used to calculate surface NO₂ concentrations from satellite columns is available upon request. Some features in the displayed maps were produced using The Climate Data Toolbox for MATLAB⁷⁰.

References

1. 1. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1223–1249 (2020).
2. 2. Pannullo, F. et al. Quantifying the impact of current and future concentrations of air pollutants on respiratory disease risk in England. *Environ. Health* **16**, 29 (2017).
3. 3. Tao, Y., Mi, S., Zhou, S., Wang, S. & Xie, X. Air pollution and hospital admissions for respiratory diseases in Lanzhou, China. *Environ. Pollut.* **185**, 196–201 (2014).
4. 4. Zeng, W. et al. Association between NO₂ cumulative exposure and influenza prevalence in mountainous regions: a case study from southwest China. *Environ. Res.* **189**, 109926 (2020).
5. 5. Anenberg, S. C. et al. Estimates of the global burden of ambient PM_{2.5}, ozone, and NO₂ on asthma incidence and emergency room visits. *Environ. Health Perspect.* **126**, 107004 (2018).
6. 6. Achakulwisut, P., Brauer, M., Hystad, P. & Anenberg, S. C. Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO₂

pollution: estimates from global datasets. *Lancet Planet. Health* **3**, e166–e178 (2019).

7. 7.

Hamra, G. B. et al. Lung cancer and exposure to nitrogen dioxide and traffic: a systematic review and meta-analysis. *Environ. Health Perspect.* **123**, 1107–1112 (2015).

8. 8.

Brook, J. R. et al. Further interpretation of the acute effect of nitrogen dioxide observed in Canadian time-series studies. *J. Expo. Sci. Environ. Epidemiol.* **17**, S36–S44 (2007).

9. 9.

Crouse, D. L. et al. Within-and between-city contrasts in nitrogen dioxide and mortality in 10 Canadian cities; a subset of the Canadian Census Health and Environment Cohort (CanCHEC). *J. Expo. Sci. Environ. Epidemiol.* **25**, 482–489 (2015).

10. 10.

Goldberg, D. L. et al. Disentangling the impact of the COVID-19 lockdowns on urban NO₂ from natural variability. *Geophys. Res. Lett.* **47**, e2020GL089269 (2020).

11. 11.

Biswal, A. et al. COVID-19 lockdown induced changes in NO₂ levels across India observed by multi-satellite and surface observations. *Atmos. Chem. Phys.* **21**, 5235–5251 (2021).

12. 12.

Koukouli, M.-E. et al. Sudden changes in nitrogen dioxide emissions over Greece due to lockdown after the outbreak of COVID-19. *Atmos. Chem. Phys.* **21**, 1759–1774 (2021).

13. 13.

Field, R. D., Hickman, J. E., Geogdzhayev, I. V., Tsigaridis, K. & Bauer, S. E. Changes in satellite retrievals of atmospheric composition over eastern China

during the 2020 COVID-19 lockdowns. Preprint at <https://doi.org/10.5194/acp-2020-567> (2020).

14. 14.

Bauwens, M. et al. Impact of coronavirus outbreak on NO₂ pollution assessed using TROPOMI and OMI observations. *Geophys. Res. Lett.* **47**, e2020GL087978 (2020).

15. 15.

Liu, F. et al. Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19. *Sci. Adv.* **6**, eabc2992 (2020).

16. 16.

Prunet, P., Lezeaux, O., Camy-Peyret, C. & Thevenon, H. Analysis of the NO₂ tropospheric product from S5P TROPOMI for monitoring pollution at city scale. *City Environ. Interact.* **8**, 100051 (2020).

17. 17.

Shi, X. & Brasseur, G. P. The response in air quality to the reduction of Chinese economic activities during the COVID-19 Outbreak. *Geophys. Res. Lett.* **47**, e2020GL088070 (2020).

18. 18.

Ropkins, K. & Tate, J. E. Early observations on the impact of the COVID-19 lockdown on air quality trends across the UK. *Sci. Total Environ.* **754**, 142374 (2021).

19. 19.

Fu, F., Purvis-Roberts, K. L. & Williams, B. Impact of the COVID-19 pandemic lockdown on air pollution in 20 major cities around the world. *Atmosphere* **11**, 1189 (2020).

20. 20.

Venter, Z. S., Aunan, K., Chowdhury, S. & Lelieveld, J. COVID-19 lockdowns cause global air pollution declines. *Proc. Natl Acad. Sci.* **117**, 18984–18990 (2020).

21. 21.

Levy, I., Mihele, C., Lu, G., Narayan, J. & Brook, J. R. Evaluating multipollutant exposure and urban air quality: pollutant interrelationships, neighborhood variability, and nitrogen dioxide as a proxy pollutant. *Environ. Health Perspect.* **122**, 65–72 (2014).

22. 22.

Shi, Z. et al. Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. *Sci. Adv.* **7**, eabd6696 (2021).

23. 23.

Liu, Q. et al. Spatiotemporal changes in global nitrogen dioxide emission due to COVID-19 mitigation policies. *Sci. Total Environ.* **776**, 146027 (2021).

24. 24.

Lamsal, L. N. et al. Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. *J. Geophys. Res.* **113**, D16308 (2008).

25. 25.

Geddes, J. A., Martin, R. V., Boys, B. L. & van Donkelaar, A. Long-term trends worldwide in ambient NO₂ concentrations inferred from satellite observations. *Environ. Health Perspect.* **124**, 281–289 (2016).

26. 26.

Gu, J. et al. Ground-level NO₂ concentrations over China inferred from the satellite OMI and CMAQ model simulations. *Remote Sens.* **9**, 519 (2017).

27. 27.

Cooper, M. J., Martin, R. V., McLinden, C. A. & Brook, J. R. Inferring ground-level nitrogen dioxide concentrations at fine spatial resolution applied to the TROPOMI satellite instrument. *Environ. Res. Lett.* **15**, 104013 (2020).

28. 28.

Levelt, P. F. et al. The Ozone Monitoring Instrument: overview of 14 years in space. *Atmos. Chem. Phys.* **18**, 5699–5745 (2018).

29. 29.

Levelt, P. F. et al. The Ozone Monitoring Instrument. *IEEE Trans. Geosci. Remote Sens.* **44**, 1093–1100 (2006).

30. 30.

Veefkind, J. P. et al. TROPOMI on the ESA Sentinel-5 Precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens. Environ.* **120**, 70–83 (2012).

31. 31.

Goldberg, D. L., Anenberg, S., Mohegh, A., Lu, Z. & Streets, D. G. TROPOMI NO₂ in the United States: a detailed look at the annual averages, weekly cycles, effects of temperature, and correlation with PM2.5. Preprint at <https://doi.org/10.1002/essoar.10503422.1> (2020).

32. 32.

Dix, B. et al. Nitrogen oxide emissions from US oil and gas production: recent trends and source attribution. *Geophys. Res. Lett.* **47**, e2019GL085866 (2020).

33. 33.

Schenkeveld, V. M. E. et al. In-flight performance of the Ozone Monitoring Instrument. *Atmos. Meas. Tech.* **10**, 1957–1986 (2017).

34. 34.

Gkatzelis, G. I. et al. The global impacts of COVID-19 lockdowns on urban air pollution: a critical review and recommendations. *Elem. Sci. Anthr.* **9**, 00176 (2021).

35. 35.

Benítez-García, S.-E., Kanda, I., Wakamatsu, S., Okazaki, Y. & Kawano, M. Analysis of criteria air pollutant trends in three Mexican metropolitan areas. *Atmosphere* **5**, 806–829 (2014).

36. 36.

Duncan, B. N. et al. A space-based, high-resolution view of notable changes in urban NO_x pollution around the world (2005–2014). *J. Geophys. Res.* **121**, 976–

996 (2016).

37. 37.

Bari, M. & Kindzierski, W. B. Fifteen-year trends in criteria air pollutants in oil sands communities of Alberta, Canada. *Environ. Int.* **74**, 200–208 (2015).

38. 38.

Zheng, B. et al. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmos. Chem. Phys.* **18**, 14095–14111 (2018).

39. 39.

Georgoulias, A. K., van der, A. R. J., Stammes, P., Boersma, K. F. & Eskes, H. J. Trends and trend reversal detection in 2 decades of tropospheric NO₂ satellite observations. *Atmos. Chem. Phys.* **19**, 6269–6294 (2019).

40. 40.

Krotkov, N. A. et al. Aura OMI observations of regional SO₂ and NO₂ pollution changes from 2005 to 2015. *Atmos. Chem. Phys.* **16**, 4605–4629 (2016).

41. 41.

Hilboll, A., Richter, A. & Burrows, J. P. NO₂ pollution over India observed from space – the impact of rapid economic growth, and a recent decline. Preprint <https://doi.org/10.5194/acp-2017-101> (2017).

42. 42.

Zhang, R. et al. Comparing OMI-based and EPA AQS in situ NO₂ trends: towards understanding surface NO_x emission changes. *Atmos. Meas. Tech.* **11**, 3955–3967 (2018).

43. 43.

Lin, N., Wang, Y., Zhang, Y. & Yang, K. A large decline of tropospheric NO₂ in China observed from space by SNPP OMPS. *Sci. Total Environ.* **675**, 337–342 (2019).

44. 44.

Barkley, M. P. et al. OMI air-quality monitoring over the Middle East. *Atmos. Chem. Phys.* **17**, 4687–4709 (2017).

45. 45.

Vohra, K. et al. Long-term trends in air quality in major cities in the UK and India: a view from space. *Atmos. Chem. Phys.* **21**, 6275–6296 (2021).

46. 46.

Kerr, G. H., Goldberg, D. L. & Anenberg, S. C. COVID-19 pandemic reveals persistent disparities in nitrogen dioxide pollution. *Proc. Natl Acad. Sci.* **118**, e2022409118 (2021).

47. 47.

Le, T. et al. Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China. *Science* **369**, 702–706 (2020).

48. 48.

Chen, L.-W. A., Chien, L.-C., Li, Y. & Lin, G. Nonuniform impacts of COVID-19 lockdown on air quality over the United States. *Sci. Total Environ.* **745**, 141105 (2020).

49. 49.

Hammer, M. S. et al. Effects of COVID-19 lockdowns on fine particulate matter concentrations. *Sci. Adv.* **7**, eabg7670 (2021).

50. 50.

Keller, C. A. et al. Global impact of COVID-19 restrictions on the surface concentrations of nitrogen dioxide and ozone. *Atmos. Phys. Chem.* **21**, 3555–3592 (2021).

51. 51.

Lamsal, L. N. et al. Ozone Monitoring Instrument (OMI) Aura nitrogen dioxide standard product version 4.0 with improved surface and cloud treatments. *Atmos. Meas. Tech.* **14**, 455–479 (2021).

52. 52.

van Geffen, J. et al. S5P TROPOMI NO₂ slant column retrieval: method, stability, uncertainties and comparisons with OMI. *Atmos. Meas. Tech.* **13**, 1315–1335 (2020).

53. 53.

Folkert Boersma, K. et al. Improving algorithms and uncertainty estimates for satellite NO₂ retrievals: results from the quality assurance for the essential climate variables (QA4ECV) project. *Atmos. Meas. Tech.* **11**, 6651–6678 (2018).

54. 54.

Goldberg, D. L. et al. Enhanced capabilities of TROPOMI NO₂: estimating NO_x from North American cities and power plants. *Environ. Sci. Technol.* **53**, 12594–12601 (2019).

55. 55.

Spurr, R. *Area-weighting Tessellation For Nadir-Viewing Spectrometers*. Internal Technical Note (Harvard-Smithsonian Center for Astrophysics, 2003).

56. 56.

Zhu, L. et al. Formaldehyde (HCHO) as a hazardous air pollutant: mapping surface air concentrations from satellite and inferring cancer risks in the United States. *Environ. Sci. Technol.* **51**, 5650–5657 (2017).

57. 57.

Bey, I. et al. Global modeling of tropospheric chemistry with assimilated meteorology: model description and evaluation. *J. Geophys. Res. Atmos.* **106**, 23073–23095 (2001).

58. 58.

Park, R. J., Jacob, D. J., Field, B. D., Yantosca, R. M. & Chin, M. Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: implications for policy. *J. Geophys. Res. Atmos.* **109**, D15204 (2004).

59. 59.

Rienecker, M. M. et al. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Clim.* **24**, 3624–3648 (2011).

60. 60.
Hammer, M. S. et al. Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environ. Sci. Technol.* **54**, 7879–7890 (2020).
61. 61.
Gatz, D. F. & Smith, L. The standard error of a weighted mean concentration—I. Bootstrapping vs other methods. *Atmos. Environ.* **29**, 1185–1193 (1995).
62. 62.
Chimot, J., Vlemmix, T., Veefkind, J. P., de Haan, J. F. & Levelt, P. F. Impact of aerosols on the OMI tropospheric NO₂ retrievals over industrialized regions: how accurate is the aerosol correction of cloud-free scenes via a simple cloud model? *Atmos. Meas. Tech.* **9**, 359–382 (2016).
63. 63.
Lin, J.-T. et al. Retrieving tropospheric nitrogen dioxide from the Ozone Monitoring Instrument: effects of aerosols, surface reflectance anisotropy, and vertical profile of nitrogen dioxide. *Atmos. Chem. Phys.* **14**, 1441–1461 (2014).
64. 64.
Cooper, M. J., Martin, R. V., Hammer, M. S. & McLinden, C. A. An observation-based correction for aerosol effects on nitrogen dioxide column retrievals using the Absorbing Aerosol Index. *Geophys. Res. Lett.* **46**, 8442–8452 (2019).
65. 65.
Verhoelst, T. et al. Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO₂ measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks. *Atmos. Meas. Tech.* **14**, 481–510 (2021).
66. 66.
Laughner, J. L., Zare, A. & Cohen, R. C. Effects of daily meteorology on the interpretation of space-based remote sensing of NO₂. *Atmos. Chem. Phys.* **16**, 15247–15264 (2016).
67. 67.

Liu, S. et al. An improved air mass factor calculation for nitrogen dioxide measurements from the Global Ozone Monitoring Experiment-2 (GOME-2). *Atmos. Meas. Tech.* **13**, 755–787 (2020).

68. 68.

Judd, L. M. et al. Evaluating the impact of spatial resolution on tropospheric NO₂ column comparisons within urban areas using high-resolution airborne data. *Atmos. Meas. Tech.* **12**, 6091–6111 (2019).

69. 69.

Kharol, S. K. et al. Assessment of the magnitude and recent trends in satellite-derived ground-level nitrogen dioxide over North America. *Atmos. Environ.* **118**, 236–245 (2015).

70. 70.

Greene, C. A. et al. The Climate Data Toolbox for MATLAB. *Geochem. Geophys. Geosyst.* **20**, 3774–3781 (2015).

Acknowledgements

This research was supported by Environment and Climate Change Canada and by the Canadian Urban Environmental Health Research Consortium. R.V.M. acknowledges support from NASA grants 80NSSC21K1343 and 80NSSC21K0508. We thank the OMI instrument team, and the OMI and TROPOMI teams for making NO₂ data publicly available.

Author information

Affiliations

1. Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada

Matthew J. Cooper, Randall V. Martin & Melanie S. Hammer

2. Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis, St. Louis, MO, USA

Matthew J. Cooper, Randall V. Martin & Melanie S. Hammer

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

Randall V. Martin

4. Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

Pieter F. Levelt & Pepijn Veefkind

5. University of Technology Delft, Delft, Netherlands

Pieter F. Levelt

6. National Center for Atmospheric Research, Boulder, CO, USA

Pieter F. Levelt

7. Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, Netherlands

Pepijn Veefkind

8. NASA Goddard Space Flight Center, Greenbelt, MD, USA

Lok N. Lamsal & Nickolay A. Krotkov

9. Universities Space Research Association, Columbia, MD, USA

Lok N. Lamsal

10. Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

Jeffrey R. Brook

11. Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada

Jeffrey R. Brook

12. Environment and Climate Change Canada, Toronto, Ontario, Canada

Chris A. McLinden

Contributions

M.J.C. and R.V.M. designed the study. M.J.C. performed the analysis. M.S.H. performed GEOS-Chem model simulations and developed the PM_{2.5} data used here. P.F.L. and P.V. developed and provided the TROPOMI NO₂ data used here. L.N.L. and N.A.K. developed and provided the OMI NO₂ data used here. M.J.C. prepared the manuscript with contributions from R.V.M., M.S.H., P.F.L., P.V., L.N.L., N.A.K., J.R.B. and C.A.M.

Corresponding author

Correspondence to [Matthew J. Cooper](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Additional information

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

Supplementary information

[Supplementary Information](#)

This file contains Supplementary Methods, Supplementary Table 1, Supplementary Figures 1–20, and additional references.

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

Cooper, M.J., Martin, R.V., Hammer, M.S. *et al.* Global fine-scale changes in ambient NO₂ during COVID-19 lockdowns. *Nature* **601**, 380–387 (2022).

<https://doi.org/10.1038/s41586-021-04229-0>

- Received: 09 February 2021
- Accepted: 11 November 2021
- Published: 19 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04229-0>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

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

- Article
- Open Access
- [Published: 20 October 2021](#)

Evidence for European presence in the Americas in ad 1021

- [Margot Kuitens](#) [ORCID: orcid.org/0000-0002-8803-2650¹](#),
- [Birgitta L. Wallace²](#),
- [Charles Lindsay²](#),
- [Andrea Scifo](#) [ORCID: orcid.org/0000-0002-7174-3966¹](#),
- [Petra Doeve](#) [ORCID: orcid.org/0000-0002-8322-2068^{3,4}](#),
- [Kevin Jenkins²](#),
- [Susanne Lindauer](#) [ORCID: orcid.org/0000-0001-5363-2755⁵](#),
- [Pınar Erdil](#) [ORCID: orcid.org/0000-0001-7463-6034¹](#),
- [Paul M. Ledger^{6,7}](#),
- [Véronique Forbes](#) [ORCID: orcid.org/0000-0002-1302-7603⁶](#),
- [Caroline Vermeeren⁸](#),
- [Ronny Friedrich](#) [ORCID: orcid.org/0000-0001-5199-1957⁵](#) &
- [Michael W. Dee](#) [ORCID: orcid.org/0000-0002-3116-453X¹](#)

[Nature](#) volume 601, pages 388–391 (2022)

- 180k Accesses
- 4 Citations
- 3053 Altmetric
- [Metrics details](#)

Subjects

- [Archaeology](#)
- [Mass spectrometry](#)
- [Plant physiology](#)

Abstract

Transatlantic exploration took place centuries before the crossing of Columbus. Physical evidence for early European presence in the Americas can be found in Newfoundland, Canada^{1,2}. However, it has thus far not been possible to determine when this activity took place^{3,4,5}. Here we provide evidence that the Vikings were present in Newfoundland in ad 1021. We overcome the imprecision of previous age estimates by making use of the cosmic-ray-induced upsurge in atmospheric radiocarbon concentrations in ad 993 (ref. ⁶). Our new date lays down a marker for European cognisance of the Americas, and represents the first known point at which humans encircled the globe. It also provides a definitive tie point for future research into the initial consequences of transatlantic activity, such as the transference of knowledge, and the potential exchange of genetic information, biota and pathologies^{7,8}.

[Download PDF](#)

Main

The Vikings (or Norse) were the first Europeans to cross the Atlantic⁹. However, the only confirmed Norse site in the Americas is L'Anse aux Meadows, Newfoundland^{9,10,11,12} (Extended Data Figs. ¹ and ²). Extensive field campaigns have been conducted at this UNESCO (United Nations Educational, Scientific, and Cultural Organization) World Heritage Site, and much knowledge has been gained about the settlement and its contemporary environment^{2,13,14,15} (Supplementary Note ¹). Evidence has also revealed that L'Anse aux Meadows was a base camp from which other locations, including regions further south, were explored¹⁵.

The received paradigm is that the Norse settlement dates to the close of the first millennium⁹; however, the precise age of the site has never been scientifically established. Most previous estimates have been based on stylistic analysis of the architectural remains and a handful of artefacts, as well as interpretations of the Icelandic sagas, oral histories that were only written down centuries later^{2,16} (Supplementary Note ²). Radiocarbon (¹⁴C) analysis has been attempted at the site, but has not proved especially informative^{3,17,18}. More than 150 ¹⁴C dates have been obtained, of which 55 relate to the Norse occupation¹⁹. However, the calibrated age ranges provided by these samples extend across and beyond the entire Viking Age (ad 793–1066) (Fig. ¹ and Extended Data Fig. ³). This is in contrast with the archaeological evidence and interpretations of the sagas. The latter offer differing scenarios for the frequency and duration of Norse activity in the Americas, but both

the archaeological and written records are consistent with a very brief occupation (Supplementary Note 3 and Extended Data Fig. 4). The unfavourable spread in the ^{14}C dates is down to the limitations of this chronometric technique in the 1960s and 1970s when most of these dates were obtained. Such impediments included far greater measurement uncertainty and restrictive sample size requirements. Furthermore, many of these samples were subject to an unknown amount of inbuilt age. The term inbuilt age refers to the difference in time between the contextual age of the sample and the time at which the organism died (returned by ^{14}C analysis), which can potentially reach hundreds of years. This offset was also sometimes inappropriately incorporated into summary estimates³.

Fig. 1: Date ranges obtained from our wiggle matches in comparison with legacy ^{14}C data.

 figure 1

a, b, Averaged probability density functions for different sample types (Extended Data Fig. 3, Supplementary Note 5 and Supplementary Data 1). **a**, Samples susceptible to inbuilt age. Light blue, whale bone ($n = 1$, uncorrected for marine reservoir effect); red, wood ($n = 17$); brown, burnt wood ($n = 7$); black, charcoal ($n = 22$). **b**, Short-lived samples. Light green, turf or sod from the Norse buildings ($n = 4$); olive, outermost rings and twigs from Norse-modified wood ($n = 4$). **c**, Wiggle-matched probability density functions for the last growth ring of each wood item. Dark green, 4A 59 E3-1; navy, 4A 68 J4-6; orange, 4A 68 E2-2.

[Source data](#)

Cosmic radiation events as absolute time markers

In our study, we use an advanced chronometric approach to anchor Norse activity in the Americas to a precise point in time. Exact-year ^{14}C results can be achieved by high-precision accelerator mass spectrometry (AMS) in combination with distinct features in the atmospheric ^{14}C record^{20,21,22}. Measurements on known-age (dendrochronological) tree rings show that ^{14}C production usually fluctuates by less than 2‰ per year²³. However, such time series have also revealed that production of the isotope rapidly increased in the years ad 775 and ad 993 by about 12‰ (which manifests as a decrease of about 100 ^{14}C yr)²⁴ and about 9‰ (about 70 ^{14}C yr)⁶, respectively. These sudden increases were caused by cosmic radiation events, and appear synchronously in dendrochronological records all around the world^{25,26,27,28,29}. By uncovering these features in tree-ring samples of unknown age, it is possible to effect precise pattern matching between such samples and reference series. In so doing, if the bark edge (or more specifically, the waney edge) is also present, it becomes possible to determine the exact felling year of the tree²⁰. Moreover, it is not necessary to have ^{14}C dates for the outermost growth rings, because once the ring that contains the ad 993 anomaly has been detected, it simply becomes a matter of counting the number of rings to the waney edge. On the basis of the state of development of the earlywood and latewood cells in the waney edge, one can even determine the precise felling season.

Precise dating of Norse activity in the Americas

Here we present 127 ^{14}C measurements, of which 115 were performed at the Centre for Isotope Research (CIO; Groningen), and 12 at the Curt-Engelhorn-Center Archaeometry (CEZA; Mannheim). The samples consisted of 83 individual tree rings from a total of 4 wooden items with find numbers 4A 59 E3-1, 4A 68 E2-2, 4A 68 J4-6 and 4A 70 B5-14 (Extended Data Fig. 5, Supplementary Note 4 and Supplementary Data 2). Unfortunately, the last item is excluded from the remainder of our analysis because it spans only nine years and does not include the ad 993 anomaly and therefore cannot be precisely dated (Supplementary Data 2). Anatomical characteristics such as different numbers of growth rings, varying growth-ring widths and the presence-absence of features such as missing rings show that wood items 4A 59 E3-1, 4A 68 E2-2 and 4A 68 J4-6 come from different trees. Furthermore, they comprise at least two different species, specifically fir, possibly balsam fir (*Abies cf. balsamea*), and juniper/thuja (*Juniperus/Thuja* type; Extended Data Fig. 6). In addition, the waney edge could be identified in all cases.

The items were found at the locations shown on the site map in Extended Data Fig. 2. The association of these pieces with the Norse is based on detailed research previously

conducted by Parks Canada. The determining factors were their location within the Norse deposit and the fact that they had all been modified by metal tools, evident from their characteristically clean, low angle-in cuts³⁰. Such implements were not manufactured by the Indigenous inhabitants of the area at the time³⁰ (Supplementary Note 4).

Our individual ^{14}C results are consistently better than $\pm 2.5\text{\%}$ (1σ), with some averaged results better than $\pm 1.5\text{\%}$ (about 12 ^{14}C yr). Our corpus of replicated measurements is consistent with statistical expectation, and no statistically significant offset (5.1 ± 7.9 ^{14}C yr, 1σ) was evident between the two ^{14}C facilities involved (Supplementary Data 2).

Two steps are used to determine the exact cutting year of each piece of wood. First, the range of possible dates for the waney edges is obtained by standard ^{14}C wiggle matching against the Northern Hemisphere calibration curve, IntCal20 (ref. 23). Here we use the D_Sequence function in the software OxCal (ref. 31) to match the full ^{14}C time-series for each item. The resultant 95% probability (2σ) ranges for the waney edges all lie between ad 1019 and ad 1024 (Fig. 1c). This indicates that the ad 993 anomaly should be present in each of the wood pieces 26 to 31 years before they were cut. In our numbering system, this corresponds to rings -31 to -26 , where the waney edge is assigned to be 0, the penultimate ring is assigned to be -1 , and so forth.

A second step is then used to determine the exact cutting year of each item. This process hinges on identifying the precise ring in which the ad 993 anomaly is found, and hence the precise date of the waney edge. For this purpose, we use the Classical χ^2 approach^{20,32} to match the ^{14}C data from the six rings (-31 to -26) most likely to contain the ad 993 anomaly against a second Northern Hemisphere reference (henceforth B2018)²⁸. This dataset is preferred because the ad 993 anomaly is less distinct in the smoothed IntCal20 curve (Fig. 2). The six-ring subsets are compared with B2018 such that χ^2 becomes minimal for the cutting date of each item. The matches are conducted over a range for each waney edge of ad 1016–1026 (Fig. 2a).

Fig. 2: Exact date matches obtained from the χ^2 tests.

 **figure 2**

The wood items are identified as follows: 4A 59 E3-1 (dark green); 4A 68 J4-6 (navy); 4A 68 E2-2 (orange). **a**, Outputs of the χ^2 test against B2018 (ref. [28](#); d.f. = 5, critical value = 11.07, 95% probability), where the gold cross marks the year of best fit for the waney edge. **b**, All of the ^{14}C data from 4A 59 E3-1 ($n = 12$, 1σ), 4A 68 J4-6 ($n = 35$, 1σ) and 4A 68 E2-2 ($n = 29$, 1σ) superimposed on IntCal20 (light blue, 1σ). Inset: detail of the ^{14}C results (error bars omitted for legibility) for growth rings –31 to –26 against B2018 (grey, 1σ)[28](#) and IntCal20 (light blue).

Source data

The optimal χ^2 value for goodness-of-fit for the waney edge in all three cases is ad 1021 (Fig. [2a](#)). While other solutions pass the χ^2 test at 95% probability (ad 1022 for 4A 59 E3-1; ad 1022 for 4A 68 E2-2; ad 1019, ad 1020 and ad 1022 for 4A 68 J4-6), the ideal positioning for the precipitous drop in ^{14}C years in each case is when ring –29 corresponds to ad 992 (inset of Fig. [2b](#)). Furthermore, the formation of a small band of earlywood cells in 4A 68 J4-6 indicates a felling season in spring (Extended Data Fig. [7a](#)). The felling season of 4A 68 E2-2 is summer/autumn (Extended Data Fig. [7b](#)). Past polyethylene glycol ([Methods](#)) consolidation hinders determination of the felling season of 4A 59 E3-1.

Our result of ad 1021 for the cutting year constitutes the only secure calendar date for the presence of Europeans across the Atlantic before the voyages of Columbus.

Moreover, the fact that our results, on three different trees, converge on the same year is notable and unexpected. This coincidence strongly suggests Norse activity at L'Anse aux Meadows in ad 1021. Further evidence reinforces this conclusion. First, the modifications are extremely unlikely to have taken place before this year, because the globally observed sudden decrease in ^{14}C values is evident in ring –29. Second, the probability that the items would have been modified at a later stage is also negligible. This is largely because of the fact that they all had their waney edges preserved. This layer would almost certainly have been stripped off during water transport, so the possibility of driftwood is effectively discounted³³. Further, the Norse would have had no need to reclaim deadwood because fresh wood was abundant in the region at the time¹³. Finally, if it were scavenged material, the probability that all three items would exhibit precisely the same amount of inbuilt age would be vanishingly small.

The Icelandic sagas suggest that the Norse engaged in cultural exchanges with the Indigenous groups of North America³⁴. If these encounters indeed occurred, they may have had inadvertent outcomes, such as pathogen transmission⁷, the introduction of foreign flora and fauna species, or even the exchange of human genetic information. Recent data from the Norse Greenlandic population, however, show no evidence of the last of these⁸. It is a matter for future research how the year ad 1021 relates to overall transatlantic activity by the Norse. Nonetheless, our findings provide a chronological anchor for further investigations into the consequences of their westernmost expansion.

Conclusions

We provide evidence that the Norse were active on the North American continent in the year ad 1021. This date offers a secure juncture for late Viking chronology. More importantly, it acts as a new point-of-reference for European cognisance of the Americas, and the earliest known year by which human migration had encircled the planet. In addition, our research demonstrates the potential of the ad 993 anomaly in atmospheric ^{14}C concentrations for pinpointing the ages of past migrations and cultural interactions. Together with other cosmic-ray events, this distinctive feature will allow for the exact dating of many other archaeological and environmental contexts.

Methods

Sampling

After careful examination of the transversal and radial sections of the wood, and ring counting, individual samples were collected under a microscope for annual-ring measurement using a steel blade, following the standard procedure for cleaving tree rings. Sample extraction started at the waney edge. For each wood item, the sample of the waney edge was given the number 0, the second-to-last ring was given the number -1, and so forth.

Sample preparation and measurement

The tree-ring samples were cut into small fragments again using a steel blade. All of the wood samples were chemically pretreated and analysed at CIO, Groningen. For independent control, 12 of the samples were also chemically pretreated and analysed at CEZA, Mannheim. CEZA and CIO recently took part in a multi-laboratory intercomparison exercise to ensure the effectiveness of their pretreatment protocols in which tree-ring samples of unknown age were pretreated to α -cellulose and then analysed for ^{14}C concentration by AMS³⁵.

Procedures at CIO, University of Groningen

The first step involves pretreating the samples to α -cellulose, the most rigid and immobile fraction of the wood³⁶. The method has previously been described in full³⁷. In brief, the wet chemistry involves a series of strong solutions of acid–base–acid and an oxidant, with rinses to neutrality using deionized and ultrapure water after each step. The samples are then either freeze-dried or air-dried at room temperature for 72 h. To eliminate the additive polyethylene glycol (PEG), which was present in all wood items except 4A 68 E2-2, the aqueous pretreatment is preceded by placement of the samples in ultrapure water at 80 °C for 36 h. This latter step builds on past studies of this contaminant^{38,39,40}. In cases where the starting weight was <20 mg, and the wood was not treated with PEG, the holocellulose protocol used at CIO was deemed sufficient³⁷.

Aliquots (about 5 mg, where possible) of the (alpha-)cellulosic product are weighed into tin capsules for combustion in an elemental analyser (IsotopeCube, Elementar). A small amount of the $\text{CO}_2(\text{g})$ released is directed into an isotope ratio mass spectrometer (Isoprime 100) for determination of the stable isotope ratios of C and N, but the majority is cryogenically trapped into Pyrex rigs and reduced to graphite under a stoichiometric excess of $\text{H}_2(\text{g})$ over an $\text{Fe}(\text{s})$ catalyst. The graphite (about 2 mg) is subsequently pressed into $\text{Al}(\text{s})$ cathodes for measurement by AMS (MICADAS, Ionplus). The data were refined using BATS 4.0 and stored in FileMaker Pro 14.6.0. For quality control purposes, full pretreatment and radioisotope measurements were concurrently conducted on known-age standards (for example, tree-ring material from

ad 1503, UK) and background wood (Pleistocene deposit Kitzbühel, Austria). Community-wide isotope ratio mass spectrometry and AMS standards (for example, National Institute of Standards and Technology oxalic acid II, Merck caffeine, and International Atomic Energy Agency C7 and C8) were used to validate the isotope measurements.

Procedures at CEZA, Mannheim

Samples MAMS-45877–45879 and MAMS-47884–47886 are pretreated as holocellulose and are pretreated using the acid–base–acid method (acid/base/acid, HCl/NaOH/HCl) followed by bleaching with NaClO₂ to extract the cellulose⁴¹. The second batch of samples (MAMS-50444–50449) is pretreated as alpha-cellulose following the protocol used by CIO described above. PEG contamination is removed in the same way as at CIO by washing in hot ultrapure water. The cellulose is combusted to CO₂ in an elemental analyser. CO₂ is then converted catalytically to graphite. ¹⁴C is analysed in-house using an AMS instrument of the MICADAS type. The isotopic ratios (¹⁴C/¹²C of samples, calibration standard oxalic acid II), blanks and control standards are measured simultaneously in the AMS. ¹⁴C ages are normalized to $\delta^{13}\text{C} = -25\text{\textperthousand}$ (ref. ⁴²), where $\delta^{13}\text{C} = (((^{13}\text{C}/^{12}\text{C})_{\text{sample}}/(^{13}\text{C}/^{12}\text{C})^{\text{standard}}) - 1) \times 1,000$.

Models in the program OxCal

All models employ OxCal 4.4 and use its standard Metropolis–Hastings Markov chain Monte Carlo algorithm and default priors³¹. The code for these models is provided in Supplementary Note 5 and in the repository <https://github.com/mwdee/LAM1021>.

Averaging

Averages are produced for each sample type using the Sum function in OxCal 4.4. In each case, all of the relevant ¹⁴C dates are included in bounded phases. The main prior information used by this model is that each date is assumed to be part of a defined group³¹.

Wiggle matching

¹⁴C data for each beam are wiggle matched against the IntCal20 calibration curve in OxCal 4.4 using its D_Sequence function³¹. All models show high convergence and run to completion.

Pattern matching using the χ^2 test

The measured ^{14}C concentrations of tree-ring samples are matched to a reference curve through the classical statistical method of the χ^2 test^{[20,22](#)}, using the following χ^2 function:

$$\text{\$\$}\{{\{X\}}^2\}_{\{(x)\}}=\text{\mathop{\sum }}\limits_{i=1}^n\frac{\{(\{R\}_i-C(x-r_i))\}^2}{\{\rm{\delta }R_i\}^2+\{\rm{\delta }C(x-r_i)\}^2}\text{\$\$}$$

Here $R_i \pm \delta R_i$ are the measured ^{14}C dates of the sample; $C(x) \pm \delta C(x)$ are the ^{14}C concentrations of the reference curve for the year $(x) - r_i$, where r_i are the tree-ring numbers of the samples analysed; and x is a trial age for the waney edge. Measured dates are matched to the reference data (that is, either higher or lower) in such a way that the χ^2 becomes minimal for a certain value of x , which is the best estimate for the felling date of the tree^{[20](#)}. To match the event accurately, a reference dataset is needed that has single-year resolution. We use B2018 as this reference, which combines many annual ^{14}C results for the years relevant to this study^{[28](#)}. The pattern-matching analyses are predominantly carried out using Python 3 in Jupyter Notebook 6.3.0. The results on each of the wood items studied are shown in Fig. 2.

Wood taxonomy

From the three main fragments of wood (4A 59 E3-1, 4A 68 E2-2 and 4A 68 J4-6), thin sections are prepared under a stereomicroscope with magnifications of up to $50\times$. They are cut in three directions (transverse, radial and tangential). As the wood was dry, the sections had to be soaked in soapy water to get rid of air bubbles and to be able to see the diagnostic anatomical features. The slides are examined under a transmitted light microscope with magnifications up to $\times 400$ and identified with the help of relevant literature^{[43,44,45](#)}. The three samples do not have any vessels, and therefore must be softwood from conifer species. The most important characteristics for identification are the lack of resin canals, the height of the rays (on average much lower in 4A 68 J4-6 than in the other two samples) and the type, number and distribution of the crossfield pits. Also, presence/absence of axial parenchyma, the shape of the ray cells in crossfields, the pitting in side walls and end walls of the ray cells, and the geographical provenance are taken into account. As wood sample 4A 68 J4-6 is compression wood (reaction wood on the lower side of branches and leaning stems), the distinction between cupressoid and taxoidoid pits cannot be made. The identification for this sample is therefore uncertain with juniper and thuja as possible candidates (*Juniperus/Thuja* type). The other two samples are identified with confidence as fir (*Abies*). Within this genus further identification is impossible, but balsam fir (*A. balsamea*), a very common North American species, would be a good match.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

All of the data that support the findings of this study are available in the main text or [Supplementary Information](#). [Source data](#) are provided with this paper.

Code availability

The codes of the OxCal models are provided in the [Supplementary Information](#) and in the repository <https://github.com/mwdee/LAM1021>.

References

1. 1.

Ingstad, H. & Ingstad, A. S. *The Viking Discovery of America: The Excavations of a Norse Settlement at L'Anse aux Meadows, Newfoundland* (Breakwater Books, 2000).

2. 2.

Wallace, B. L. in *Contact, Continuity, and Collapse: the Norse Colonization of the North Atlantic* (ed. Barrett, J.) 207–238 (Brepols, 2003).

3. 3.

Nydal, R. A critical review of radiocarbon dating of a Norse settlement at L'Anse aux Meadows, Newfoundland Canada. *Radiocarbon* **31**, 976–985 (1989).

4. 4.

Ledger, P. M., Girdland-Flink, L. & Forbes, V. New horizons at L'Anse aux Meadows. *Proc. Natl Acad. Sci. USA* **116**, 15341–15343 (2019).

5. 5.

Dee, M. W. & Kuitens, M. Duration of activity inestimable due to imprecision of the data. *Proc. Natl Acad. Sci. USA* **116**, 22907 (2019).

6. 6.

Miyake, F., Masuda, K. & Nakamura, T. Another rapid event in the carbon-14 content of tree rings. *Nat. Commun.* **4**, 1748 (2013).

7. 7.

Mühlemann, B. et al. Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. *Science* **369**, 6502 (2020).

8. 8.

Margaryan, A. et al. Population genomics of the Viking world. *Nature* **585**, 390–396 (2020).

9. 9.

Wallace, B. L. The Norse in Newfoundland: L'Anse aux Meadows and Vinland. *Newfoundl. Labrador Stud.* **19**, 50–43 (2003).

10. 10.

Ingstad, H. *The Discovery Norse House-Sites in North America* (Harper & Row, 1966).

11. 11.

Lindsay, C. S. Was L'Anse aux Meadows a Norse outpost? *Can. Geogr. J.* **94**, 36–43 (1977).

12. 12.

Ingstad, A. S. & Ingstad, H. *The Norse Discovery of America* Vols I and II (Univ. Oslo Press, 1986).

13. 13.

Davis, A. M., McAndrews, J. H. & Wallace, B. L. Paleoenvironment and the archaeological record at the L'Anse aux Meadows site, Newfoundland. *Geoarchaeology* **3**, 53–64 (1988).

14. 14.

Ogilvie, A. E., Barlow, L. K. & Jennings, A. E. North Atlantic climate c.AD 1000: millennial reflections on the Viking discoveries of Iceland, Greenland and North America. *Weather* **55**, 34–45 (2000).

15. 15.

Wallace, B. L. L'Anse aux Meadows, Leif Eriksson's home in Vinland. *J. North Atl.* Special Vol. 2, 114–125 (2009).

16. 16.

Smiley, J. *The Sagas of the Icelanders* (Penguin, 2005).

17. 17.

Martindale, A. et al. *Canadian Archaeological Radiocarbon Database (CARD 2.1)* (accessed 13 April 2021) (2016).

18. 18.

Kristensen, T. J. & Curtis, J. E. Late Holocene hunter-gatherers at L'Anse aux Meadows and the dynamics of bird and mammal hunting in Newfoundland. *Arctic Anthropol.* **49**, 68–87 (2012).

19. 19.

Wallace, B. L. in *Archaeology in America: An Encyclopedia* (eds Cordell, L. S. et al.) 78–83 (ABC-CLIO, 2009).

20. 20.

Wacker, L. et al. Radiocarbon dating to a single year by means of rapid atmospheric ^{14}C changes. *Radiocarbon* **56**, 573–579 (2016).

21. 21.

Oppenheimer, C. et al. Multi-proxy dating the 'Millennium Eruption' of Changbaishan to late 946 CE. *Quat. Sci. Rev.* **158**, 164–171 (2017).

22. 22.

Kuitems, M. et al. Radiocarbon-based approach capable of subannual precision resolves the origins of the site of Por-Bajin. *Proc. Natl Acad. Sci. USA* **117**, 14038–14041 (2020).

23. 23.

Reimer, P. et al. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).

24. 24.

Miyake, F., Nagaya, K., Masuda, K. & Nakamura, T. A signature of cosmic-ray increase in AD 774–775 from tree rings in Japan. *Nature* **486**, 240–242 (2012).

25. 25.

Usoskin, I. G. et al. The AD775 cosmic event revisited: the Sun is to blame. *Astron. Astrophys.* **552**, L3 (2013).

26. 26.

Jull, A. T. et al. Excursions in the ^{14}C record at A.D. 774–775 in tree rings from Russia and America. *Geophys. Res. Lett.* **41**, 3004–3010 (2014).

27. 27.

Güttler, D. et al. Rapid increase in cosmogenic ^{14}C in AD 775 measured in New Zealand kauri trees indicates short-lived increase in ^{14}C production spanning both hemispheres. *Earth Planet. Sci. Lett.* **411**, 290–297 (2015).

28. 28.

Büntgen, U. et al. Tree rings reveal globally coherent signature of cosmogenic radiocarbon events in 774 and 993 CE. *Nat. Commun.* **9**, 3605 (2018).

29. 29.

Scifo, A. et al. Radiocarbon production events and their potential relationship with the Schwabe cycle. *Sci. Rep.* **9**, 17056 (2019).

30. 30.

Wallace, B. L. *Westward to Vinland: the Saga of L'Anse aux Meadows* (Historic Sites Association of Newfoundland and Labrador, 2012).

31. 31.

Bronk Ramsey, C. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).

32. 32.

Bronk Ramsey, C., van der Plicht, J. & Weninger, B. ‘Wiggle matching’ radiocarbon dates. *Radiocarbon* **43**, 381–389 (2001).

33. 33.

Mooney, D. E. A. ‘North Atlantic island signature’ of timber exploitation: evidence from wooden artefact assemblages from Viking Age and Medieval Iceland. *J. Archaeol. Sci. Rep.* **7**, 280–289 (2016).

34. 34.

Odess, D., Loring, S. & Fitzhugh W. W. in *Vikings: the North Atlantic Saga* (eds Fitzhugh, W. W. & Ward, E. I.) (Smithsonian Institution Press, 2000).

35. 35.

Wacker, L. et al. Findings from an in-depth annual tree-ring radiocarbon intercomparison. *Radiocarbon* **62**, 873–882 (2020).

36. 36.

Loader, N. J., Robertson, I. & McCarroll, D. Comparison of stable carbon isotope ratios in the whole wood, cellulose and lignin of oak tree-rings. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **196**, 395–407 (2003).

37. 37.

Dee, M. W. et al. Radiocarbon dating at Groningen: new and updated chemical pretreatment procedures. *Radiocarbon* **62**, 63–74 (2020).

38. 38.

Brock, F. et al. Testing the effectiveness of protocols for removal of common conservation treatments for radiocarbon dating on dating. *Radiocarbon* **60**, 35–50 (2018).

39. 39.

Bruhn, F., Duhr, A., Grootes, P. M., Mintrop, A., Nadeau, M.-J. Chemical removal of conservation substances by ‘Soxhlet’-type extraction. *Radiocarbon* **43**, 229–237 (2001).

40. 40.

Ensing, B. et al. On the origin of the extremely different solubilities of polyethers. *Nat. Commun.* **10**, 2893 (2019).

41. 41.

Friedrich, R. et al. Annual ^{14}C tree-ring data around 400 AD: mid- and high-latitude records. *Radiocarbon* **61**, 1305–1316 (2019).

42. 42.

Stuiver, M. & Polach, H. A. Discussion reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).

43. 43.

Schweingruber F. H. *Anatomy of European Woods* (Bern and Stuttgart, 1990).

44. 44.

Wheeler, E. A. InsideWood - a web resource for hardwood anatomy. *International Association of Wood Anatomists Journal* **32**, 199–211 (2011).

45. 45.

IAWA Committee. IAWA list of microscopic features for softwood identification. *IAWA J.* **25**, 1–70 (2004).

Acknowledgements

This work was funded by the European Research Council (grant 714679, ECHOES). M.K., A.S., P.E. and M.W.D. were supported by this grant. We thank Parks Canada for providing samples; the CIO staff, especially S. W. L. Palstra, D. van Zonneveld, R. Linker, S. de Bruin, R. A. Schellekens, P. Wietzes-Land, D. Paul, H. A. J. Meijer, J. J. Spriensma, H. G. Jansen, A. Th. Aerts-Bijma and A. C. Neocleous; and R. Doeve, E. van Hees, A. J. Huizinga, B. J. S. Pope and J. Higdon for their help and support.

Author information

Affiliations

1. Centre for Isotope Research, University of Groningen, Groningen, the Netherlands

Margot Kuitems, Andrea Scifo, Pınar Erdil & Michael W. Dee

2. Parks Canada Agency, Government of Canada, Dartmouth, Nova Scotia, Canada

Birgitta L. Wallace, Charles Lindsay & Kevin Jenkins

3. Laboratory for Dendrochronology at BAAC, 's-Hertogenbosch, the Netherlands

Petra Doeve

4. Cultural Heritage Agency of The Netherlands, Amersfoort, the Netherlands

Petra Doeve

5. Curt-Engelhorn-Center Archaeometry, Mannheim, Germany

Susanne Lindauer & Ronny Friedrich

6. Department of Archaeology, Queens College, Memorial University of Newfoundland, St Johns, Newfoundland, Canada

Paul M. Ledger & Véronique Forbes

7. Department of Geography, Memorial University of Newfoundland, St Johns, Newfoundland, Canada

Paul M. Ledger

8. BIAX Consult, Zaandam, the Netherlands

Caroline Vermeeren

Contributions

M.W.D. conceived the idea, directed the research and co-wrote the paper; M.K. helped to design the research, conducted most of it and co-wrote the paper; B.L.W. was principal advisor on archaeology and sagas; C.L. advised on archaeology; A.S. mainly performed the χ^2 analyses; P.D. advised on tree-ring anatomy; K.J. took samples; S.L. conducted pretreatments (Mannheim); P.E. conducted pretreatments (Groningen);

P.M.L. and V.F. advised on archaeology and palaeoecology; C.V. analysed wood taxonomy; R.F. oversaw AMS analyses (Mannheim). All co-authors contributed to the final draft of the manuscript.

Corresponding authors

Correspondence to [Margot Kuitens](#) or [Michael W. Dee](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks James Barrett, Dagfinn Skre, Lukas Wacker and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

Extended data figures and tables

[Extended Data Fig. 1 North Atlantic regions explored by the Norse.](#)

LAM lies on the Northern Peninsula of Newfoundland. The map shows the main settlements on Greenland from where the Norse embarked, and the regions they named Helluland, Markland and Vinland. Map: R. Klaarenbeek.

[Extended Data Fig. 2 Schematic overview of the site \(after Wallace 2003\)2 and origin of our samples.](#)

Indicated are the contours of different Norse structures (A–J) and the locations (brown) at which the wood items were found that are used in the current study.

[Extended Data Fig. 3 The 55 legacy \$^{14}\text{C}\$ dates on Norse contexts at LAM.](#)

Samples susceptible to inbuilt age: light blue, whale bone ($n = 1$, uncorrected for Marine Reservoir Effect); red, wood ($n = 17$); brown, burnt wood ($n = 7$); black, charcoal ($n = 22$). Short-lived samples: light green, turf or sod samples from the walls of the Norse buildings ($n = 4$); olive, outermost rings and twigs from Norse-modified wood ($n = 4$). See Supplementary Data 1.

Extended Data Fig. 4 Overview of the number and order of the different voyages by the Norse to the Americas based on the information from the Sagas.

Indicated for each voyage are the expedition leader (EL), the duration (D), the number of attending crew and the number of ships. Top, summary of the information from the *Saga of the Greenlanders*, which indicates that the number of winters spent at Vinland is seven. Given the short sailing seasons and the impossibility of making round trips between Greenland and Vinland in one year, the time between the first arrival of the Norse at Vinland and their ultimate return is estimated to be about thirteen years; bottom, summary of the information from the *Saga of Erik the Red*, with the estimated minimum time between the first arrival of the Norse at Vinland and their ultimate return, amounting to about three years.

Extended Data Fig. 5 Pictures of the wood items studied.

White X indicates the location from where samples were taken. The black bars represent 5 cm. Top left, 4A 59 E3-1; top right, 4A 68 E2-2; bottom left, 4A 68 J4-6; bottom right, 4A 70 B5-14. Photos: M. Kuitens.

Extended Data Fig. 6 Microscope pictures of the thin slices from the wood samples studied.

The white bars represent 0.05 mm, the black bars 0.1 mm. From left to right: radial, tangential and transversal sections of respectively: top, 4A 68 J4-6; middle, 4A 68 E2-2; bottom, 4A 59 E3-1. Photos: M. van Waijen, BIAX Consult.

Extended Data Fig. 7 Microscopic depiction of the felling season of the waney edge.

The black bars represent 1 mm. ew = early wood, which is formed during the first stage of the growth year; lw = late wood, which is formed at the end of the growth season. **a**, Wood item 4A 68 E2-2; **b**, Wood item 4A 68 J4-6. Photos: P. Doeve.

Supplementary information

Supplementary Information

This file contains Supplementary Notes 1–5 and references. (1) L'Anse aux Meadows; (2) Dating of the site; (3) Length of occupation; (4) Sample materials; (5) Codes.

Reporting Summary

Peer Review File

Source data

Source Data Fig. 1

Source Data Fig. 2

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

Kuitems, M., Wallace, B.L., Lindsay, C. *et al.* Evidence for European presence in the Americas in ad 1021. *Nature* **601**, 388–391 (2022). <https://doi.org/10.1038/s41586-021-03972-8>

- Received: 21 May 2021

- Accepted: 31 August 2021
- Published: 20 October 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-03972-8>

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

Further reading

- [**A radiocarbon revolution sheds light on the Vikings**](#)
 - James H. Barrett

Nature (2022)
- [**Single-year radiocarbon dating anchors Viking Age trade cycles in time**](#)
 - Bente Philippsen
 - Claus Feveile
 - Søren M. Sindbæk

Nature (2022)

[**Vikings were living in North America exactly a thousand years ago**](#)

- Shaminibundell
- Dan Fox

Nature Video 22 Oct 2021

A radiocarbon revolution sheds light on the Vikings

- James H. Barrett

News & Views 22 Dec 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-03972-8>

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

- Article
- [Published: 22 December 2021](#)

Single-year radiocarbon dating anchors Viking Age trade cycles in time

- [Bente Philippsen](#) ORCID: [orcid.org/0000-0001-5732-3570^{1,2}](https://orcid.org/0000-0001-5732-3570),
- [Claus Feveile³](#),
- [Jesper Olsen](#) ORCID: [orcid.org/0000-0002-4445-5520^{1,2}](https://orcid.org/0000-0002-4445-5520) &
- [Søren M. Sindbæk²](#)

Nature volume **601**, pages 392–396 (2022)

- 2920 Accesses
- 1 Citations
- 169 Altmetric
- [Metrics details](#)

Subjects

- [Applied physics](#)
- [Archaeology](#)
- [Carbon cycle](#)
- [History](#)

Abstract

Recent discoveries of rapid changes in the atmospheric ^{14}C concentration linked to solar particle events have spurred the construction of new radiocarbon annual calibration datasets^{[1](#),[2](#),[3](#),[4](#),[5](#),[6](#),[7](#),[8](#),[9](#),[10](#),[11](#),[12](#),[13](#)}. With these datasets, radiocarbon dating becomes relevant for urban sites, which require dates at higher resolution than previous calibration datasets could offer. Here we use a single-year radiocarbon calibration curve to anchor the archaeological stratigraphy of a Viking Age trade centre in time. We present absolutely dated evidence for artefact finds charting the expansion of long-distance trade from as far away as Arctic Norway and the Middle East, which we linked to the beginning of the Viking Age at ad 790 ± 10 . The methods developed here enable human interactions and cultural, climatic and environmental changes to be compared in archaeological stratigraphies worldwide.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.
VAT will be added later in the checkout.
Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Radiocarbon dating results of annual tree rings, compared to the international calibration curve IntCal13.

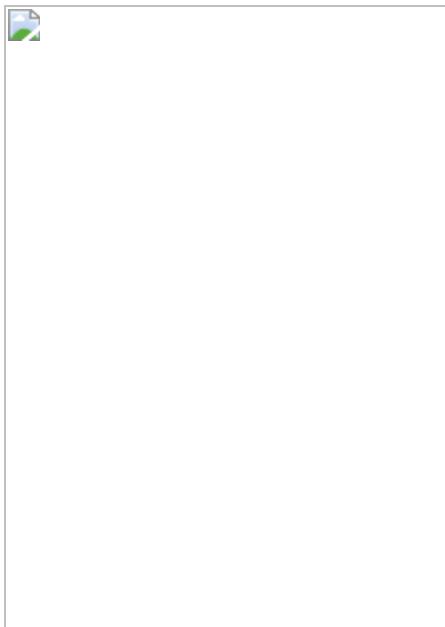


Fig. 2: Viking Age artefacts and trade connections through time.

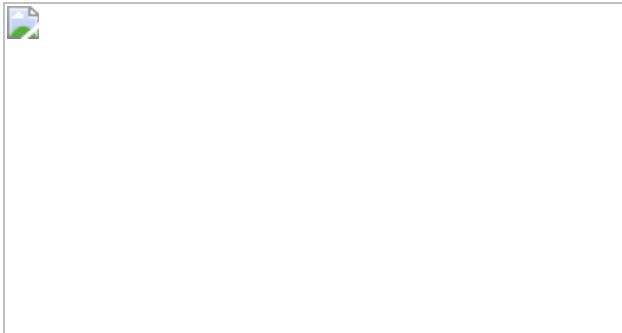
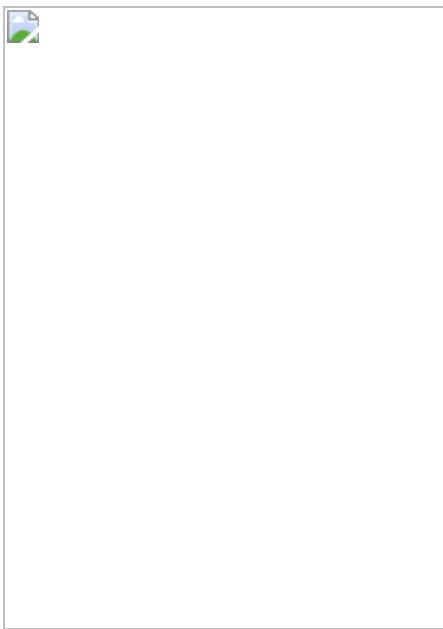


Fig. 3: Radiocarbon dates, artefact distributions and climate during phase 9 in Ribe.



Data availability

All data are available in the Article or the [Supplementary Information](#). Additional information about the site and excavation can be found at <https://projects.au.dk/northernemporium/>, and additional artefact photographs can be found at <http://sol.sydvestjyskemuseer.dk/> using the search term ‘SJM 3’.

Code availability

All code is available in the [Supplementary Information](#).

References

1. 1.

Reimer, P. J. et al. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 kcal BP). *Radiocarbon* **62**, 725–757 (2020).

2. 2.

Büntgen, U. et al. Tree rings reveal globally coherent signature of cosmogenic radiocarbon events in 774 and 993 CE. *Nat. Commun.* **9**, 3605 (2018).

3. 3.

Dee, M. et al. Supernovae and single-year anomalies in the atmospheric radiocarbon record. *Radiocarbon* **59**, 293–302 (2016).

4. 4.

Fogtmann-Schulz, A. Cosmic ray event in 994 C.E. recorded in radiocarbon from Danish oak. *Geophys. Res. Lett.* **44**, 8621–8628 (2017).

5. 5.

Jull, A. J. T. et al. More rapid ^{14}C excursions in the tree-ring record: a record of different kind of solar activity at about 800 BC? *Radiocarbon* **60**, 1237–1248 (2018).

6. 6.

Wang, F. Y. et al. A rapid cosmic-ray increase in BC 3372–3371 from ancient buried tree rings in China. *Nat. Commun.* **8**, 1487 (2017).

7. 7.

Mekhaldi, F. et al. Multiradionuclide evidence for the solar origin of the cosmic-ray events of AD 774/5 and 993/4. *Nat. Commun.* **6**, 8611 (2015).

8. 8.

Park, J. et al. Relationship between solar activity and ^{14}C peaks in AD 775, AD 994, and 660 BC. *Radiocarbon* **59**, 1147–1156 (2017).

9. 9.

Miyake, F. et al. Large ^{14}C excursion in 5480 BC indicates an abnormal sun in the mid-Holocene. *Proc. Natl Acad. Sci. USA* **114**, 881–884 (2017).

10. 10.

Dee, M. W. & Pope, B. J. S. Anchoring historical sequences using a new source of astro-chronological tie-points. *Proc. R. Soc. A* **472**, 20160263 (2016).

11. 11.

Wacker, L. et al. Radiocarbon dating to a single year by means of rapid atmospheric ^{14}C changes. *Radiocarbon* **56**, 573–579 (2014).

12. 12.

Kuitems, M. et al. Radiocarbon-based approach capable of subannual precision resolves the origins of the site of Por-Bajin. *Proc. Natl Acad. Sci. USA* **117**, 14038–14041 (2020).

13. 13.

Miyake, F. et al. A signature of cosmic-ray increase in AD 774–775 from tree rings in Japan. *Nature* **486**, 240–242 (2012).

14. 14.

Büntgen, U. et al. 2500 years of European climate variability and human susceptibility. *Science* **331**, 578–582 (2011).

15. 15.

Cook, E. R. et al. Old World megadroughts and pluvials during the Common Era. *Sci. Adv.* **1**, e1500561 (2015).

16. 16.

Misra, P., Tandon, S. K. & Sinha, R. Holocene climate records from lake sediments in India: assessment of coherence across climate zones. *Earth Sci. Rev.* **190**, 370–397 (2019).

17. 17.

Denniston, R. F. & Luetscher, M. Speleothems as high-resolution paleoflood archives. *Quat. Sci. Rev.* **170**, 1–13 (2017).

18. 18.

Dahl-Jensen, D. et al. Eemian interglacial reconstructed from a Greenland folded ice core. *Nature* **493**, 489–494 (2013).

19. 19.

Margaryan, A. et al. Population genomics of the Viking world. *Nature* **585**, 390–396 (2020).

20. 20.

Hansen, V. *The Year 1000. When Explorers Connected the World – And Globalization Began* (Scribner, 2020).

21. 21.

Hodges, R. & Whitehouse, D. *Mohammed, Charlemagne and the Origins of Europe. The Pirenne Thesis in the Light of Archaeology* (Duckworth, 1983).

22. 22.

Noonan, T. S. *The Islamic World, Russia and the Vikings, 750–900. The Numismatic Evidence* (Routledge, 1998).

23. 23.

McCormick, M. *Origins of the European Economy: Communications and Commerce AD 300–900* (Cambridge Univ. Press, 2001).

24. 24.

Jankowiak, M. in *Viking-Age Trade: Silver, Slaves and Gotland* (eds Gruszczyński, J. et al.) (Routledge, 2020).

25. 25.

Barrett, J. H. What caused the Viking Age? *Antiquity* **82**, 671–685 (2008).

26. 26.

Hodges, R. *Dark Age Economics: A New Audit* (Bloomsbury Academic, 2012).

27. 27.

Wickham, C. *Framing the Early Middle Ages: Europe and the Mediterranean, 400–800* (Oxford Univ. Press, 2005).

28. 28.

Baug, I. et al. The beginning of the Viking Age in the West. *J. Marit. Archaeol.* **14**, 43–80 (2019).

29. 29.

Schrijver, C. J. et al. Estimating the frequency of extremely energetic solar events, based on solar, stellar, lunar, and terrestrial records. *J.*

Geophys. Res. Space Phys. **117** (2012).

30. 30.

Baroni, M. et al. Volcanic and solar activity, and atmospheric circulation influences on cosmogenic ^{10}Be fallout at Vostok and Concordia (Antarctica) over the last 60 years. *Geochim. Cosmochim. Acta* **75**, 7132–7145 (2011).

31. 31.

Stuiver, M. & Braziunas, T. F. Sun, ocean, climate and atmospheric $^{14}\text{CO}_2$: an evaluation of causal and spectral relationships. *The Holocene* **3**, 289–305 (1993).

32. 32.

Kuitems, M. et al. Evidence for European presence in the Americas in AD 1021. *Nature* <https://doi.org/10.1038/s41586-021-03972-8> (2021).

33. 33.

Croix, S. et al. Single context, metacontext, and high definition archaeology: Integrating new standards of stratigraphic excavation and recording. *J. Archaeol. Method Theory* **26**, 1591–1631 (2019).

34. 34.

Yang, J. & Ren, P. BFDA: a MATLAB toolbox for Bayesian functional data analysis. *J. Stat. Softw.* **89**, 21 (2019).

35. 35.

Buck, C. E. et al. Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity* **65**, 808–821 (1991).

36. 36.

Bronk Ramsey, C. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).

37. 37.

Callmer, J. in *Glass beads – Cultural History, Technology, Experiment and Analogy. Proceedings of the Nordic Glass Bead Seminar 16th–18th October 1992 Studies in Technology and Culture 2*. Lejre (eds Rasmussen, M. et al.) 49–54 (1995).

38. 38.

Sindbæk, S. M. in *Urban Network Evolutions: Towards a high-definition archaeology* (eds Raja, R. and Sindbæk, S.) 161–166 (Aarhus Univ. Press, 2018).

39. 39.

Ashby, S., Coutu, A. & Sindbæk, S. Urban networks and arctic outlands: craft specialists and reindeer antler in Viking towns. *Eur. J. Archaeol.* **18**, 679–704 (2015).

40. 40.

Luterbacher, J. et al. European summer temperatures since Roman times. *Environ. Res. Lett.* **11**, 024001 (2016).

41. 41.

Kerr, T. R., Swindles, G. T. & Plunkett, G. Making hay while the sun shines? Socio-economic change, cereal production and climatic deterioration in Early Medieval Ireland. *J. Archaeolog. Sci.* **36**, 2868–2874 (2009).

42. 42.

Sukhodolov, T. et al. Atmospheric impacts of the strongest known solar particle storm of 775 AD. *Sci. Rep.* **7**, 45257 (2017).

43. 43.

Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).

44. 44.

Hall, R. *Exploring the World of the Vikings* (Thames and Hudson, 2007).

45. 45.

Croix, S. et al. Single context, metacontext, and high definition archaeology: integrating new standards of stratigraphic excavation and recording. *J. Archaeol. Method Theory* **26**, 1591–1631 (2019).

46. 46.

Tyers, I. *DENDRO for Windows Program Guide* ARCUS Report Vol. 500 (Univ. of Sheffield, 1999).

47. 47.

Baillie, M. & Pilcher, J. A simple cross-dating program for tree-ring research. *Tree-Ring Bull.* **33**, 7–14 (1973).

48. 48.

Kudsk, S. G. K., et al. What is the carbon origin of early-wood? *Radiocarbon* **60**, 1457–1464 (2018).

49. 49.

McDonald, L., D. Chivall, Miles, D. & Bronk Ramsey, C. Seasonal variations in the ^{14}C content of tree rings: influences on radiocarbon calibration and single-year curve construction. *Radiocarbon* **61**, 185–194 (2018).

50. 50.

Loer, N. J., Robertson, I., Barker, A. C., Switsur, V. R. & Waterhouse, J. S. An improved technique for the batch processing of small wholewood samples to α -cellulose. *Chem. Geol.* **136**, 313–317 (1997).

51. 51.

Southon, J. R. & Magana, A. L., A comparison of cellulose extraction and ABA pretreatment methods for AMS C-14 dating of ancient wood. *Radiocarbon* **52**, 1371–1379 (2010).

52. 52.

Kudsk, S. G. K. et al. New single-year radiocarbon measurements based on Danish oak covering the periods AD 692–790 and 966–1057. *Radiocarbon* **62**, 969–987 (2019).

53. 53.

Vogel, J. S., Southon, J. R., Nelson, D. E. & Brown, T. A., Performance of catalytically condensed carbon for use in accelerator mass spectrometry. *Nucl. Instrum. Methods Phys. Res. B* **5**, 289–293 (1984).

54. 54.

Olsen, J., Tikhomirov, D., Grosen, C., Heinemeier, J. & Klein, M.. Radiocarbon analysis on the new AARAMS 1MV Tandetron. *Radiocarbon* **59**, 905–913 (2016).

55. 55.

Stuiver, M. & Polach, H. A. Discussion. Reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).

56. 56.

Longin, R. New method of collagen extraction for radiocarbon dating. *Nature* **230**, 241–242 (1971).

57. 57.

Brown, T. A., Nelson, D. E., Vogel, J. C. & Sounthor, J. R. Improved collagen extraction by improved Longin method. *Radiocarbon* **30**, 171–177 (1988).

58. 58.

Jørkov, M. L. S., Heinemeier, J. & Lynnerup, N. Evaluating bone collagen extraction methods for stable isotope analysis in dietary studies. *J. Archaeolog. Sci.* **34**, 1824–1829 (2007).

Acknowledgements

We thank the excavators of the site, whose work formed the basis for this study: H. Brinch Christiansen, M. Knudsen, S. Qvistgaard and M. Søvsø from the Museum of Southwest Jutland, and S. Croix and P. Deckers from Aarhus University. The large numbers of radiocarbon dates would not have been possible without the support of the staff and PhD students of the Aarhus AMS Centre: A. Fogtmann-Schulz, C. Grosen, H. Jakobsen, M. Kanstrup, S. Kudsk, M. Sand Kalaee and A. B. Valbøl Jensen. This study was funded by the Carlsberg Foundation Semper Ardens grant no. CF16-0008 (Northern Emporium project) and the Danish National Research Foundation grant no. DNRF119 – Centre of Excellence for Urban Network Evolutions (UrbNet). Brødrene Hartmanns Fond (grant no. application A34514) and Grosserer P.L. Jørgensens Mindefond supported the tree-ring measurements.

Author information

Affiliations

1. Aarhus AMS Centre (AARAMS), Department of Physics and Astronomy, Aarhus University, Aarhus, Denmark

Bente Philippsen & Jesper Olsen

2. Centre for Urban Network Evolutions (UrbNet), Aarhus University,
Højbjerg, Denmark

Bente Philippsen, Jesper Olsen & Søren M. Sindbæk

3. Museum of Southwest Jutland, Ribe, Denmark

Claus Feveile

Contributions

J.O. and S.M.S. conceptualized the study. B.P., C.F., J.O. and S.M.S. were responsible for data curation. B.P. and J.O. carried out the formal analysis. S.M.S., J.O. and B.P. were responsible for funding acquisition. B.P., C.F., J.O. and S.M.S. carried out the investigations. B.P., C.F., J.O. and S.M.S. were responsible for the methodology. S.M.S. and C.F. administered the project. B.P. and J.O. were responsible for visualization. B.P., S.M.S. and J.O. wrote the original draft. B.P. and C.F. reviewed and edited the manuscript.

Corresponding author

Correspondence to [Bente Philippsen](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks James Barrett, Paula Reimer and Dagfinn Skre 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 Finds relating to the manufacture of wound glass beads.

a) glass vessel cullet; b) mosaic tesserae; c) splints from batches of re-melted, coloured glass; d) production debris, including droplets; e) cane ends with plier marks; f) canes for applied trail ornaments; g) ‘Ribe type’ beads; h) ‘wasp type’ beads. Photos: Museum of Southwest Jutland. High-resolution images of all finds are available at <http://sol.sydvestjyskemuseer.dk/> using the search term “SJM 3”.

Extended Data Fig. 2 Finds relating to non-ferrous metalworking.

a) crucible sherds; b) fragments of clay casting moulds; c) mould fragment with impressions of a cast brooch; d) mould fragment for Berdal-type brooches. Photos: Museum of Southwest Jutland. High-resolution images of all finds are available at <http://sol.sydvestjyskemuseer.dk/> using the search term “SJM 3”.

Extended Data Fig. 3 Common types of imported beads found in Ribe.

a) small segmented ‘gold-foil’ bead; b) segmented ‘gold -foil’ bead; c) segmented colourless ‘silver-foil’ bead; d) segmented blue bead; e) segmented blue metal-foil bead; f) green faceted bead; g) mosaic eye beads; h) cut tubular beads; i) blown metal-foil beads. Photos: Museum of Southwest Jutland. High-resolution images of all finds are available at <http://sol.sydvestjyskemuseer.dk/> using the search term “SJM 3”.

Extended Data Fig. 4 Common types of imports from the Rhine area.

a) fragments of Mayen basalt quern stones; b) fragment of Badorf ware pottery; c) fragment of Reliefband amphora; d) fragment of Tating ware pitcher. Photos: Museum of Southwest Jutland. High-resolution images of all finds are available at <http://sol.sydvestjyskemuseer.dk/> using the search term “SJM 3”.

Extended Data Fig. 5 Common types of imports from the Scandinavian Peninsula.

a) fragments of (or blanks for) whetstones made from dark and light schist; b) casting mould made of soap stone; c) sherds of soap stone vessels. Photos: Museum of Southwest Jutland. High-resolution images of all finds are available at <http://sol.sydvestjyskemuseer.dk/> using the search term “SJM 3”.

Extended Data Fig. 6 Examples of sceatta coins types.

a) Wodan/Monster (W/M) – obverse and reverse; b) Continental Runic – obverse and reverse; c) Porcupine – obverse and reverse. Photos: Museum of Southwest Jutland. High-resolution images of all finds are available at <http://sol.sydvestjyskemuseer.dk/> using the search term “SJM 3”.

Supplementary information

Supplementary Information

Methods and Results descriptions (excavation and artefact chronology; radiocarbon dating); thirteen tables detailing artefact distribution per phases on this project’s excavation (SJM 3) and on the neighbouring excavation ASR 9 Posthuset from 1990 to 1991; one table with ^{14}C ages for all tree-ring samples; two tables with detailed information about samples for radiocarbon dating and dendrochronology, as well as ^{14}C ages and

unmodelled and modelled radiocarbon dating results of these samples, calibrated with IntCal20 and the Aarhus curve; thirteen supplementary figures, of which Figs. 1 to 6 show typical artefacts from the excavations, as can be found in the Extended Data, Fig. 7 displays a map of the study area and the neighbouring excavation ASR 9 Posthuset from 1990 to 1991, and Figs. 8–13 are related to radiocarbon analyses, calibration curves and age models.

Reporting Summary

Peer Review File

Supplementary Data

The MATLAB code used in this study.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Philippsen, B., Feveile, C., Olsen, J. *et al.* Single-year radiocarbon dating anchors Viking Age trade cycles in time. *Nature* **601**, 392–396 (2022).
<https://doi.org/10.1038/s41586-021-04240-5>

- Received: 20 April 2021
- Accepted: 12 November 2021
- Published: 22 December 2021
- Issue Date: 20 January 2022

- DOI: <https://doi.org/10.1038/s41586-021-04240-5>

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

Further reading

- [**A radiocarbon revolution sheds light on the Vikings**](#)

- James H. Barrett

Nature (2022)

[**A radiocarbon revolution sheds light on the Vikings**](#)

- James H. Barrett

News & Views 22 Dec 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04240-5>

- Article
- [Published: 15 December 2021](#)

Individual human cortical progenitors can produce excitatory and inhibitory neurons

- [Ryan N. Delgado](#) [ORCID: orcid.org/0000-0002-7551-0182](#)^{1,2,3,4 na1},
- [Denise E. Allen](#) [ORCID: orcid.org/0000-0001-9205-2796](#)^{1,2,3,4 na1},
- [Matthew G. Keefe](#) [ORCID: orcid.org/0000-0002-2477-2975](#)^{1,2,3,4 na1},
- [Walter R. Mancia Leon](#)^{3,5},
- [Ryan S. Ziffra](#)^{1,2,3,4},
- [Elizabeth E. Crouch](#)^{3,6},
- [Arturo Alvarez-Buylla](#)^{3,5} &
- [Tomasz J. Nowakowski](#) [ORCID: orcid.org/0000-0003-2345-4964](#)^{1,2,3,4,5,7}

[Nature](#) volume 601, pages 397–403 (2022)

- 8645 Accesses
- 1 Citations
- 118 Altmetric
- [Metrics details](#)

Subjects

- [Cell fate and cell lineage](#)
- [Developmental neurogenesis](#)

Abstract

The cerebral cortex is a cellularly complex structure comprising a rich diversity of neuronal and glial cell types. Cortical neurons can be broadly categorized into two classes—excitatory neurons that use the neurotransmitter glutamate, and inhibitory interneurons that use γ -aminobutyric acid (GABA). Previous developmental studies in rodents have led to a prevailing model in which excitatory neurons are born from progenitors located in the cortex, whereas cortical interneurons are born from a separate population of progenitors located outside the developing cortex in the ganglionic eminences^{1,2,3,4,5}. However, the developmental potential of human cortical progenitors has not been thoroughly explored. Here we show that, in addition to excitatory neurons and glia, human cortical progenitors are also capable of producing GABAergic neurons with the transcriptional characteristics and morphologies of cortical interneurons. By developing a cellular barcoding tool called ‘single-cell-RNA-sequencing-compatible tracer for identifying clonal relationships’ (STICR), we were able to carry out clonal lineage tracing of 1,912 primary human cortical progenitors from six specimens, and to capture both the transcriptional identities and the clonal relationships of their progeny. A subpopulation of cortically born GABAergic neurons was transcriptionally similar to cortical interneurons born from the caudal ganglionic eminence, and these cells were frequently related to excitatory neurons and glia. Our results show that individual human cortical progenitors can generate both excitatory neurons and cortical interneurons, providing a new framework for understanding the origins of neuronal diversity in the human cortex.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: STICR-labelled progenitors generate all three principal cortical cell types.

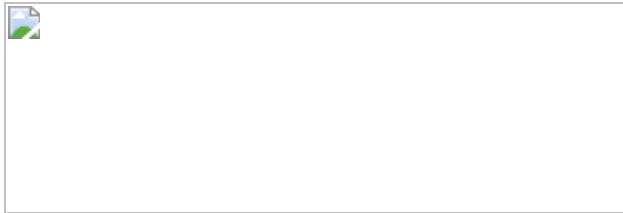


Fig. 2: Individual human cortical progenitors can generate both excitatory and inhibitory cortical neurons in vitro.

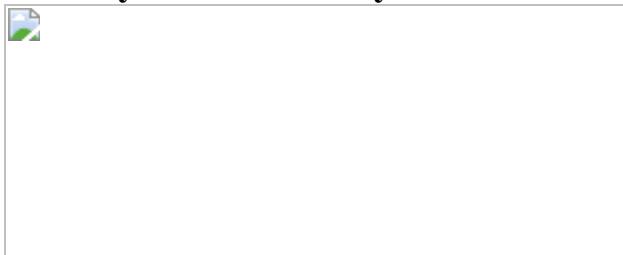


Fig. 3: Xenografted human cortical progenitors generate both excitatory and inhibitory cortical neurons in the same clone.

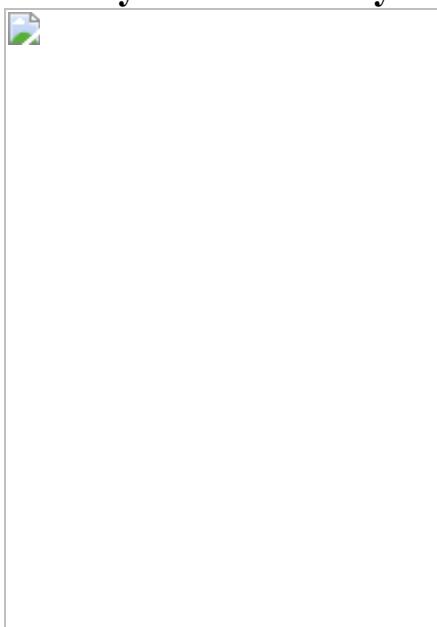


Fig. 4: Xenografted human cortical progenitors generate GABAergic inhibitory neurons that distribute across the cortical laminae.



Data availability

scRNA-seq transcriptomic data and STICR barcode data are available at the database of Genotypes and Phenotypes (dbGAP; <https://www.ncbi.nlm.nih.gov/gap/>) under accession number phs002624.v1.p1; and at the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE187875. An interactive browser of single-cell data can be found at the University of California, Santa Cruz (UCSC) cell browser⁴⁹(<https://human-cortical-lineage.cells.ucsc.edu>). Publicly available reference genomes hg38 and mm10 were used for analysis. [Source data](#) are provided with this paper.

Code availability

Custom codes used in this study are available at the following GitHub repository: <https://github.com/NOW-Lab/STICR>.

References

1. 1.

Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes. *Science* **278**, 474–476 (1997).

2. 2.

Sussel, L., Marin, O., Kimura, S. & Rubenstein, J. L. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development* **126**, 3359–3370 (1999).

3. 3.

Gorski, J. A. et al. Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).

4. 4.

Xu, Q., Tam, M. & Anderson, S. A. Fate mapping Nkx2.1-lineage cells in the mouse telencephalon. *J. Comp. Neurol.* **506**, 16–29 (2008).

5. 5.

Anderson, S. A., Marín, O., Horn, C., Jennings, K. & Rubenstein, J. L. Distinct cortical migrations from the medial and lateral ganglionic eminences. *Development* **128**, 353–363 (2001).

6. 6.

Petanjek, Z., Berger, B. & Esclapez, M. Origins of cortical GABAergic neurons in the cynomolgus monkey. *Cereb. Cortex* **19**, 249–262 (2009).

7. 7.

Letinic, K., Zoncu, R. & Rakic, P. Origin of GABAergic neurons in the human neocortex. *Nature* **417**, 645–649 (2002).

8. 8.

Hansen, D. V. et al. Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat. Neurosci.* **16**, 1576–1587 (2013).

9. 9.

Alzu’bi, A. et al. Distinct cortical and sub-cortical neurogenic domains for GABAergic interneuron precursor transcription factors NKX2.1, OLIG2 and COUP-TFII in early fetal human telencephalon. *Brain Struct. Funct.* **222**, 2309–2328 (2017).

10. 10.

Alzu’bi, A. et al. The transcription factors COUP-TFI and COUP-TFII have distinct roles in arealisation and GABAergic interneuron specification in the early human fetal telencephalon. *Cereb. Cortex* **27**, 4971–4987 (2017).

11. 11.

Kohwi, M. et al. A subpopulation of olfactory bulb GABAergic interneurons is derived from Emx1- and Dlx5/6-expressing progenitors. *J. Neurosci.* **27**, 6878–6891 (2007).

12. 12.

Young, K. M., Fogarty, M., Kessaris, N. & Richardson, W. D. Subventricular zone stem cells are heterogeneous with respect to their embryonic origins and neurogenic fates in the adult olfactory bulb. *J. Neurosci.* **27**, 8286–8296 (2007).

13. 13.

Fuentealba, L. C. et al. Embryonic origin of postnatal neural stem cells. *Cell* **161**, 1644–1655 (2015).

14. 14.

Nowakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

15. 15.

Pollen, A. A. et al. Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).

16. 16.

Bystron, I., Blakemore, C. & Rakic, P. Development of the human cerebral cortex: Boulder Committee revisited. *Nat. Rev. Neurosci.* **9**, 110–122 (2008).

17. 17.

Bandler, R. C., Mayer, C. & Fishell, G. Cortical interneuron specification: the juncture of genes, time and geometry. *Curr. Opin. Neurobiol.* **42**, 17–24 (2017).

18. 18.

Mayer, C. et al. Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).

19. 19.

Li, J. et al. Transcription factors Sp8 and Sp9 coordinately regulate olfactory bulb interneuron development. *Cereb. Cortex* **28**, 3278–3294 (2018).

20. 20.

Guo, T. et al. Dlx1/2 are central and essential components in the transcriptional code for generating olfactory bulb interneurons. *Cereb. Cortex* **29**, 4831–4849 (2019).

21. 21.

Batista-Brito, R. et al. The cell-intrinsic requirement of Sox6 for cortical interneuron development. *Neuron* **63**, 466–481 (2009).

22. 22.

Stumm, R. K. et al. CXCR4 regulates interneuron migration in the developing neocortex. *J. Neurosci.* **23**, 5123–5130 (2003).

23. 23.

López-Bendito, G. et al. Chemokine signaling controls intracortical migration and final distribution of GABAergic interneurons. *J. Neurosci.* **28**, 1613–1624 (2008).

24. 24.

Tripodi, M., Filosa, A., Armentano, M. & Studer, M. The COUP-TF nuclear receptors regulate cell migration in the mammalian basal forebrain. *Development* **131**, 6119–6129 (2004).

25. 25.

Azim, E., Jabaudon, D., Fame, R. M. & Macklis, J. D. SOX6 controls dorsal progenitor identity and interneuron diversity during neocortical development. *Nat. Neurosci.* **12**, 1238–1247 (2009).

26. 26.

Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

27. 27.

Herrera, D. G., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Adult-derived neural precursors transplanted into multiple regions in the adult brain. *Ann. Neurol.* **46**, 867–877 (1999).

28. 28.

Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).

29. 29.

Bhaduri, A. et al. An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204 (2021).

30. 30.

Lodato, M. A. et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).

31. 31.

Ludwig, L. S. et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339.e22 (2019).

32. 32.

Lareau, C. A. et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2021).

33. 33.

Paredes, M. F. et al. Extensive migration of young neurons into the infant human frontal lobe. *Science* **354**, aaf7073 (2016).

34. 34.

Bandler, R. C. et al. Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature* <https://doi.org/10.1038/s41586-021-04237-0> (2021).

35. 35.

Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).

36. 36.

Zhang, Y. et al. Cortical neural stem cell lineage progression is regulated by extrinsic signaling molecule Sonic hedgehog. *Cell Rep.* **30**, 4490–4504.e4 (2020).

37. 37.

Cai, Y., Zhang, Y., Shen, Q., Rubenstein, J. L. & Yang, Z. A subpopulation of individual neural progenitors in the mammalian dorsal pallium generates both projection neurons and interneurons in vitro. *Stem Cells* **31**, 1193–1201 (2013).

38. 38.

Daley, T. & Smith, A. D. Modeling genome coverage in single-cell sequencing. *Bioinformatics* **30**, 3159–3165 (2014).

39. 39.

Onorati, M. et al. Zika virus disrupts phospho-TBK1 localization and mitosis in human neuroepithelial stem cells and radial blia. *Cell Rep.* **16**, 2576–2592 (2016).

40. 40.

Schildge, S., Bohrer, C., Beck, K. & Schachtrup, C. Isolation and culture of mouse cortical astrocytes. *J. Vis. Exp.* **71**, 50079 (2013).

41. 41.

Crouch, E. E. & Doetsch, F. FACS isolation of endothelial cells and pericytes from mouse brain microregions. *Nat. Protocols* **13**, 738–751 (2018).

42. 42.

Bhaduri, A. et al. Outer radial blia-like cancer stem cells contribute to heterogeneity of glioblastoma. *Cell Stem Cell* **26**, 48–63.e6 (2020).

43. 43.

Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e1821 (2019).

44. 44.

Fleming, S. J., Marioni, J. C. & Babadi, M. CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. *bioRxiv* <https://doi.org/10.1101/791699> (2019).

45. 45.

Bernstein, N. J. et al. Solo: doublet identification in single-cell RNA-seq via semi-supervised deep learning. *Cell Syst.* **11**, 95–101.e5 (2020).

46. 46.

Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

47. 47.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

48. 48.

Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).

49. 49.

Speir, M. L. et al. UCSC cell browser: visualize your single-cell data. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btab503> (2021).

Acknowledgements

We thank A. Bhaduri for discussions regarding scRNA-seq analysis; B. Rabe and C. Cepko for discussions regarding viral vectors and sharing of reagents; C. Cadwell and M. Paredes for discussions regarding interneuron morphology; J. Rubenstein and R. Andersen for reading the manuscript; and M. Speir and B. Wick for data wrangling at the UCSC single-cell browser. This study was supported by the Psychiatric Cell Map Initiative Convergence Neuroscience award U01MH115747; an Innovation Award from the Broad Foundation (to T.J.N.); a New Frontiers Research Award from the Sandler Program for Breakthrough Biomedical Research (PBBR) (to T.J.N.); a National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP) award (to D.E.A.); an Autism Speaks Predoctoral Fellowship (11874 to R.S.Z.); NIH K08 NS116161 and UCSF Physician Scientist Scholars Program to EEC; and gifts from Schmidt Futures and the William K. Bowes Jr Foundation (to T.J.N.). Work in the Alvarez-Buylla laboratory is supported by National Institutes of Health (NIH) grants R01NS028478 and R01EY025174, and a gift from the John G. Bowes Research Fund. A.A.-B. is the Heather and Melanie Muss Endowed Chair and Professor of Neurological Surgery at UCSF.

Author information

Author notes

1. These authors contributed equally: Ryan N. Delgado, Denise E. Allen, Matthew G. Keefe

Affiliations

1. Department of Anatomy, University of California at San Francisco, San Francisco, CA, USA

Ryan N. Delgado, Denise E. Allen, Matthew G. Keefe, Ryan S. Ziffra & Tomasz J. Nowakowski

2. Department of Psychiatry and Behavioral Sciences, University of California at San Francisco, San Francisco, CA, USA

Ryan N. Delgado, Denise E. Allen, Matthew G. Keefe, Ryan S. Ziffra & Tomasz J. Nowakowski

3. The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California at San Francisco, San Francisco, CA, USA

Ryan N. Delgado, Denise E. Allen, Matthew G. Keefe, Walter R. Mancia Leon, Ryan S. Ziffra, Elizabeth E. Crouch, Arturo Alvarez-Buylla & Tomasz J. Nowakowski

4. Weill Institute for Neurosciences, University of California at San Francisco, San Francisco, CA, USA

Ryan N. Delgado, Denise E. Allen, Matthew G. Keefe, Ryan S. Ziffra & Tomasz J. Nowakowski

5. Department of Neurological Surgery, University of California at San Francisco, San Francisco, CA, USA

Walter R. Mancia Leon, Arturo Alvarez-Buylla & Tomasz J. Nowakowski

6. Department of Pediatrics, University of California at San Francisco, San Francisco, CA, USA

Elizabeth E. Crouch

7. Chan Zuckerberg Biohub, San Francisco, CA, USA

Tomasz J. Nowakowski

Contributions

R.N.D conceived of the project, designed and generated the STICR barcode library, designed and conducted experiments, analysed the data, and wrote the manuscript. D.E.A. helped to design experiments, conducted experiments, analysed data, and helped to write the manuscript. M.G.K helped to design experiments, conducted experiments and helped to write the manuscript. W.R.M.L carried out xenograft transplants. R.S.Z. helped to construct the STICR library. E.E.C. carried out PTPRZ1 FACS. A.A.-B. helped to supervise the research. T.J.N conceived of the project, helped to design experiments, assisted in the interpretation of data, and helped to write the manuscript.

Corresponding authors

Correspondence to [Ryan N. Delgado](#) or [Tomasz J. Nowakowski](#).

Ethics declarations

Competing interests

A.A.-B. is co-founder and on the Scientific Advisory Board of Neurona Therapeutics.

Additional information

Peer review information *Nature* thanks Zoltan Molnar and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Extended data figures and tables

[**Extended Data Fig. 1 Validation of the STICR barcode design.**](#)

a, Histogram showing pairwise hamming distances between every sequence in each STICR fragment pool. **b**, Barcode diversity extrapolations derived from sequencing a representative STICR plasmid or lentiviral library. Mean \pm 95% confidence range for each library is shown. **c**, Simulated barcode collision frequencies (mean \pm s.d.) for a range of starting cell numbers, based on the barcode diversity estimated in **b**. Barcode sampling was performed with replacement using measured proportions of barcodes within the representative plasmid and lentiviral libraries depicted in **b**. Each simulation was performed 20,000 times. Most error bars (depicting standard deviations) are not visible as they are smaller than the dots (depicting mean values). **d**, The ‘barnyard’ species-mixing experiment. **e**, Plot depicting species-specific transcript counts from barnyard experiment. Each dot depicts a single cell and the dot colour indicates whether the cell was determined to be a 3T3 cell (mouse), cortex cell (human), or mixed droplet (multiplet). **f**, Violin plots depicting the number of unique STICR barcode molecules recovered from droplets identified as either mouse, human, or multiplet. ND, not detected.

Extended Data Fig. 2 Cluster analysis of in vitro STICR data sets.

a, UMAP plots of each individual biological sample highlighted. **b**, Top marker-gene expression for each cluster. The size of each dot corresponds to the proportion of cells in the cluster that express the gene, while the colour of the dot corresponds to the average expression level per cluster. **c**, **d**, Heatmap depicting pairwise transcriptional cluster correlation of in vitro cultured cells with itself (**c**) and the 2017 Nowakowski scRNA-seq atlas (**d**)¹⁴. The principal cell-type designation is depicted next to each column and row. The dendrogram depicts hierarchical clustering distance.

Extended Data Fig. 3 Transcriptional analysis of in vitro STICR data sets.

a–f, Feature plots depicting expression of genes corresponding to cell cycle (**a**), glia (**b**), oligodendroglia (**c**), excitatory neurons (**d**), interneurons (**e**), and regional markers (**f**). **g**, Bar plot depicting the proportion of cells within

each cluster with a recovered STICR barcode. **h**, Heatmap depicting the percentage of STICR barcodes shared between biological samples. GW15 (Rep1) and all GW18 samples were labelled with the same viral stock, while GW15 (Rep2) was labelled with a different stock (see [Methods](#)). **i**, Stacked barplot depicting relative proportions of principal cell types within each sample, restricted to cells that are members of multicellular clones.

[Extended Data Fig. 4 Clonal analysis of cortical clones containing excitatory neurons.](#)

a, Histogram of excitatory neuron (EN) counts within each multicellular cortical clone. Left, clone sizes from 1–25 cells in single-cell bins. Right, clone sizes of more than 25 cells in the indicated bin sizes. **b**, Box-and-whisker plot depicting the proportion of EN cells within individual multicellular clones for each biological sample. Maxima and minima of boxes depict third and first quartiles, while box centres depict medians. Whiskers depict $1.5 \times$ the interquartile distance. Individual clone values are shown as dots. The number of clones is listed below each sample group. **c**, Ternary plots depicting the relative proportions of inhibitory neurons, excitatory neurons and all other cell types ('Other') within individual clones. **d, e**, Immunohistochemistry of in vitro cultures derived from GW15 germinal zone cells labelled with STICR. **d**, Low-magnification image to show distribution; scale bar, 25 μm . **e**, High-magnification image showing a cluster of ENs; scale bar, 250 μm .

[Source data](#)

[Extended Data Fig. 5 Clonal and transcriptional analysis of inhibitory neurons and DLX2 \$^{\pm}\$ IPCs in vitro.](#)

a, UMAP embedding and Leiden subclustering of GABAergic inhibitory neuron (IN) trajectory cells. **b**, Feature plots depicting expression of *MKI67*, *STMN2*, *CENPF* and *ERBB4*. **c**, Heatmap depicting pairwise transcriptional cluster correlation of this data set with itself. **d**, Stacked barplot depicting relative proportion of multicellular clones from each sample that comprise each IN trajectory. **e**, Stacked barplot depicting the

relative proportions of different IN trajectory cells within multicellular clones of each sample. **f**, Feature plots depicting MGE-derived cells (red) and expression of *NKX2-1*, *LHX6*, *ACKR3*, *MAF* and *PDE1A*. The enlarged insets below show IN.1 trajectory cells. **g**, Heat plot depicting differential expression of IN.2 and IN.3 marker genes in the developing human cortex, olfactory bulb/rostral migratory stream and basal ganglia. Data are derived from the Allen BrainSpan Laser Capture Microdissection database. Dendograms reflect hierarchical clustering of genes and samples while colours represent quantile-normalized *z*-scores. **h**, Paired violin plots and *in situ* hybridization (ISH) images of P60 mouse brains from the Allen Brain Atlas for select genes. The log₂ fold difference between IN.2 (olfactory-bulb-like) and IN.3 (cortical-interneuron-like) cells is depicted above each violin plot. **i**, Stacked barplots depicting relative proportions of IN.1, IN.2, IN.3, EN and glia trajectory cells within multicellular clones. The number of clones is listed below each sample. **j**, Venn diagram depicting the number of EN-containing multicellular cortical clones that also contain IN.2 and/or IN.3 cells.

Extended Data Fig. 6 Clonal and transcriptional analysis of excitatory neurons and EOMES[±] IPCs in vitro.

a, UMAP embedding and Leiden subclustering of excitatory neuron (EN) trajectory cells. **b**, **c**, Heatmap depicting pairwise transcriptional cluster correlation of subclustered EN trajectory cells with self (**b**) and with the 2017 Nowakowski developing human brain scRNA-seq atlas (**c**)¹⁴. **d**, Feature plots depicting the expression of genes corresponding to labelled subclustered EN trajectory subtypes. **e**, Stacked barplot depicting relative proportions of EN subtypes within EN trajectory cells of multicellular clones. **f**, Venn diagram depicting the number of multicellular cortical clones containing deep-like ENs, upper-like ENs, and IN.3 cells.

Extended Data Fig. 7 Characterization of human cortical progenitor xenografts at six weeks.

a, b, Representative images of transplanted human cortical cells analysed by IHC for principal cell-type markers six weeks after transplantation.

EGFP expression from STICR is in green, with NEUROD2 or GABA expression in red. Scale bars: **a**, 50 μ m; **b**, 10 μ m. **c**, Barplot depicting the proportion (mean \pm s.d.) of transplanted cells expressing principal cell-type markers as assessed by IHC. $n = 7$ sections derived from 6 xenografted mice, 3 of which were transplanted with donor cells from GW15 Rep1 and 3 with cells from GW15 Rep2. **d**, Top marker-gene expression for each cluster from xenografted cells. Sizes of dots correspond to the proportion of cells in the cluster expressing the gene, while dot colours correspond to the average expression level per cluster. **e**, UMAP embedding of xenografted cells and feature plots depicting expression of *NEUROD2*, *EOMES*, *DLX2*, *MKI67* and *GFAP*. **f**, Heatmap depicting pairwise transcriptional cluster correlation of subclustered excitatory neuron (EN) trajectory cells with the 2017 Nowakowski developing human cortex scRNA-seq atlas¹⁴. **g**, Comparison of principal cell-type quantification (mean) in transplanted cells by analysis method (IHC versus scRNA-seq) and biological replicate.

[Source data](#)

[Extended Data Fig. 8 Transcriptional analysis of excitatory and inhibitory neurons from xenografts.](#)

a, UMAP embedding and Leiden subclustering of inhibitory neuron (IN) trajectory cells from xenografts. **b**, Feature plots depicting expression of *CENPF*, *MKI67*, *ERBB4*, *NR2F1*, *NFIX*, *SP8*, *SCGN* and *KLHL35*. **c**, Heatmap depicting pairwise transcriptional cluster correlation of subclustered xenograft IN and $DXL2^+$ IPC trajectory cells with the 2017 Nowakowski developing human cortex scRNA-seq atlas¹⁴. **d**, UMAP embedding depicting cells in multicellular clones from xenograft IN subclusters 1 (salmon) and 2 (lime), integrated with an in vitro cultured STICR IN subset. **e**, UMAP embedding depicting individual interneuron trajectory cells from multicellular clones from xenograft experiments integrated with interneuron trajectory cells from in vitro cultures, split by biological replicate. Members of such clones are highlighted in red. **f**, UMAP embedding and Leiden subclustering of excitatory neuron (EN) and $EOMES^+$ IPC trajectory cells from xenografts. **g**, Heatmap depicting pairwise transcriptional cluster correlation of subclustered xenograft EN

and EOMES⁺ IPC trajectory cells with the 2017 Nowakowski developing human cortex scRNA-seq atlas¹⁴.

Extended Data Fig. 9 Analysis of PTPRZ1-sorted STICR⁺ cells in the cortex, subventricular zone, rostral migratory stream and olfactory bulb at 12 weeks.

a, Representative FACS plots depicting isolation of PTPRZ1⁺ cells from the cortical germinal zone. **b**, Representative image of transplanted human cortical cells in the cortex of a 12-week-old host mouse. EGFP expression from STICR is in green, with DAPI in blue. Scale bar, 50 µm. CC, corpus callosum. **c**, Representative images of PTPRZ1-sorted, STICR-labelled cells in the dorsolateral corner of the lateral ventricle of a 12-week-old host mouse analysed by IHC. EGFP from STICR is in green, DCX in red, and DAPI in blue. Scale bar, 100 µm. **d**, High-magnification inset of the region boxed in **c**. Scale bar, 10 µm. SVZ, subventricular zone. **e**, Representative images of PTPRZ1-sorted, STICR-labelled cells in the rostral forebrain, analysed by IHC. The rostral migratory stream (RMS) is outlined by the white box labelled **f**. GFP expression is in green, DCX expression in red, and DAPI in blue. Scale bar, 100 µm. **f**, High-magnification insets of the RMS depicted in the white box in **e**. Scale bar, 10 µm. The cell outlined by the white box is magnified below. **g**, Representative images of PTPRZ1-sorted, STICR-labelled cells that migrated from the transplantation site in the cortex to the olfactory bulb, analysed by IHC. EGFP expression is in green and DAPI in magenta. GCL, granule cell layer; MCL, mitral cell layer. Scale bar, 10 µm.

Extended Data Fig. 10 Immunohistochemistry of STICR-labelled cortical INs from xenografts at 12 weeks.

Representative images of STICR-labelled GABA⁺ cells throughout the cortical plate, analysed by IHC. EGFP from STICR is in green, GABA in red, and DAPI in blue. Same cells from Fig. [4d](#). Arrows point to STICR-labelled GABA⁺ cells. Scale bar, 10 µm.

Supplementary information

Reporting Summary

Supplemental Table 1

Table describing clone composition from *in vitro* culture experiments. Columns: Clone Barcode (“Full_VBC”), clone size (“Total Cells”), sample of origin (“Library”), as well as the number of cells within that clone assigned to principal cell types (“EN”, “IN”, “glia”, “DLX.IPC”, “EOMES.IPC”), cluster 34 (“Unknown”), the excitatory neuron subclustering cell types (“Upper”, “Deep”, “EN.Newborn”, “EN.other”), and the inhibitory neuron/DLX.IPC subclustering cell types (“DLX2.IPC”, “IN.early”, “IN.1”, “IN.2”, “IN.3”). Principal cell type designations reflect subclustering results.

Supplemental Table 2

Table describing clone composition from xenograft experiments (GW15 Reps 3, 4, and 5). Columns: Clone Barcode (“Full_VBC”), clone size (“Total Cells”), sample of origin (“Library”), as well as the number of cells within that clone assigned to principal cell types (“EN”, “IN”, “glia”, “DLX.IPC”, “EOMES.IPC”), “microglia”, “Undetermined” and the inhibitory neuron/DLX.IPC subclustering cell types (“DLX2.IPC”, “IN.early”, “IN.2”, “IN.3”). Principal cell type designations reflect subclustering results.

Supplemental Table 3

Table of STICR ssDNA oligomers used to create barcode fragment sequences.

Supplemental Table 4

Table of primers used to amplify STICR barcodes from plasmid and lentiviral libraries as well as from 10X genomics cDNA libraries.

Source data

[Source Data Fig. 4](#)

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

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

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Delgado, R.N., Allen, D.E., Keefe, M.G. *et al.* Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature* **601**, 397–403 (2022). <https://doi.org/10.1038/s41586-021-04230-7>

- Received: 15 December 2020
- Accepted: 10 November 2021
- Published: 15 December 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04230-7>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

[A shared origin for cortical excitatory neurons and interneurons](#)

Research Briefing 15 Dec 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04230-7>

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

- Article
- Open Access
- [Published: 15 December 2021](#)

Single-cell delineation of lineage and genetic identity in the mouse brain

- [Rachel C. Bandler](#)^{1,2,3} ORCID: orcid.org/0000-0001-9536-0292,
- [Ilaria Vitali](#) ORCID: orcid.org/0000-0001-9536-0292¹ ORCID: orcid.org/0000-0001-9536-0292,
- [Ryan N. Delgado](#) ORCID: orcid.org/0000-0002-7551-0182^{4,5,6} ORCID: orcid.org/0000-0002-7551-0182,
- [May C. Ho](#) ORCID: orcid.org/0000-0002-7373-2961¹,
- [Elena Dvoretskova](#)¹,
- [Josue S. Ibarra Molinas](#)¹,
- [Paul W. Frazel](#)²,
- [Maesoumeh Mohammadkhani](#)²,
- [Robert Machold](#)²,
- [Sophia Maedler](#)⁷,
- [Shane A. Liddelow](#) ORCID: orcid.org/0000-0002-0840-1437^{2,8,9},
- [Tomasz J. Nowakowski](#) ORCID: orcid.org/0000-0003-2345-4964^{4,5,6},
- [Gord Fishell](#) ORCID: orcid.org/0000-0002-9640-9278^{3,10} &
- [Christian Mayer](#) ORCID: orcid.org/0000-0003-3152-5574¹

Nature volume 601, pages 404–409 (2022)

- 20k Accesses
- 1 Citations
- 186 Altmetric
- [Metrics details](#)

Subjects

- [Cell fate and cell lineage](#)
- [Cell type diversity](#)

Abstract

During neurogenesis, mitotic progenitor cells lining the ventricles of the embryonic mouse brain undergo their final rounds of cell division, giving rise to a wide spectrum of postmitotic neurons and glia^{1,2}. The link between developmental lineage and cell-type diversity remains an open question. Here we used massively parallel tagging of progenitors to track clonal relationships and transcriptomic signatures during mouse forebrain development. We quantified clonal divergence and convergence across all major cell classes postnatally, and found diverse types of GABAergic neuron that share a common lineage. Divergence of GABAergic clones occurred during embryogenesis upon cell-cycle exit, suggesting that differentiation into subtypes is initiated as a lineage-dependent process at the progenitor cell level.

[Download PDF](#)

Main

The central nervous system consists of diverse types of neurons and glia that vary widely in morphology, physiology, connectivity and molecular markers^{3,4}. During development, molecular diversity is initially reflected in the regional expression of a narrow set of transcription factors in mitotic progenitors³. Transcriptional signatures that distinguish mature neuronal subtypes emerge only after cell-cycle exit and become more sharply defined during postnatal development^{1,2,5,6,7}. The extent to which developmental trajectories are predetermined by specified progenitor lineages during mitotic stages, or emerge through interactions with the environment later in development, remains an open question.

Although previous lineage analyses have elucidated the spatial distribution of clones, they provided little information regarding subtype identities of sister cells^{8,9,10,11,12}. More recently, breakthroughs in cellular barcoding strategies and single-cell sequencing^{13,14,15,16,17,18,19,20,21} have facilitated the recording of lineage tags and gene expression profiles in in vitro systems²², in zebrafish^{20,23,24} and in mouse embryogenesis²⁵, but have not yet been used to study neurogenesis in the mouse forebrain.

Here we combined high-throughput single-cell RNA sequencing (scRNA-seq) with massively parallel tagging of progenitors to reconstruct lineage relationships during neurogenesis of the forebrain. We focus our analysis on GABAergic neurons, which displayed a surprising degree of clonal divergence among different types of inhibitory neurons. We found that immediately after cell-cycle exit, GABAergic neurons that originated from the same mitotic progenitor diverged into different developmental

trajectories. Our findings thus revealed that differentiation into GABAergic subtypes is initiated as a lineage-dependent process at the progenitor cell level.

Capture of gene expression and lineages

To determine lineage relationships of diverse cell types in the mouse forebrain, we first implemented a lentiviral lineage barcoding method called STICR (scRNA-seq-compatible tracer for identifying clonal relationships; Fig. 1a, Extended Data Fig. 1a, see companion paper²⁶), which enables massively parallel tagging of single cells using a high-diversity lentiviral library that encodes synthetic oligonucleotide sequences (lineage barcodes). The STICR tag library was introduced via in utero injections into the lateral ventricles of mouse embryos at embryonic day 10.5 (E10.5; STICR^{E10}), E12.5 (STICR^{E12}), E13.5 (STICR^{E13}) and E14.5 (STICR^{E14}), stages that encompass the peak of neurogenesis. This resulted in labelling of mitotic progenitors along the ventricles and their daughter cells that migrated throughout the forebrain, including the cortex, basal ganglia, hippocampus and olfactory bulb (OB) (Fig. 1b, Extended Data Fig. 1b). We waited until postnatal stages when labelled cells differentiated into mature cell types, then dissociated forebrain tissue, FACS-enriched the virally infected cells by selecting for enhanced GFP (eGFP) expression, and performed scRNA-seq with the 10x Chromium System (Fig. 1a, Extended Data Fig. 1c). We analysed transcriptomes from 65,700 high-quality cells that passed filtering (see [Methods](#)). To group cells on the basis of patterns of gene expression, we performed a principal components analysis²⁷ and batch normalized the different replicates using Harmony²⁸, followed by a UMAP visualization and clustering analysis (Extended Data Fig. 1d, Supplementary Data 1), and tracked the position of clonally related cells in the transcriptomic cell-state landscape (Extended Data Fig. 1e–h). The average and maximum size of multicellular clones was larger when the lentiviral library was introduced at E10.5 than at E12.5 or E14.5, when mitotic progenitors presumably undergo fewer divisions (Fig. 1c).

Fig. 1: Clonal relationships of forebrain cell classes determined via simultaneous capture of transcriptome and lineage barcodes from single cells.

 **figure 1**

a, Schematic of STICR experimental workflow. **b**, Images of coronal brain sections injected with STICR. **c**, Quantification of the average clone size. The error bars indicate s.e.m. Each dot represents a clone. $n = 1,117$ for STICR^{E10}, $n = 169$ for STICR^{E12}, $n = 349$ for STICR^{E13} and $n = 1,407$ for STICR^{E14}. One-way analysis of

variance (ANOVA) with Tukey post-hoc test was performed; ** $P < 0.0015$, *** $P < 0.0001$. **d**, Uniform manifold approximation and projection (UMAP) plot of single cells coloured by cell classes. In the inset, a schematic of cell class origins. **e**, Heatmaps of STICR^{E10} and STICR^{E14} lineage coupling scores between pairs of cell classes, clustered by correlation distance and linkage. The light grey lines link classes across stages. **f**, Quantification of clones with 100%, more than 80%, and less than 80% of cells within a single class. **g**, Quantification of clones containing neurons, glia or a mix. **h**, Number of neurons and glial cells per clone is plotted against one another. The dotted line represents a smooth local regression; the shadow represents the 95% confidence interval.

During embryogenesis, asymmetrically dividing radial glia and transiently amplifying mitotic progenitors along the ventricular surface give rise to postmitotic neurons, astrocytes and oligodendrocytes^{3,4,29,30,31,32} (Fig. 1d). To quantify clonal relationships between cell classes, we assigned the 41 clusters into cell classes (Fig. 1d) based on the co-expression of multiple marker genes (Extended Data Fig. 2a,b, Supplementary Data 2) and counted the distribution of STICR tags across cell classes (Extended Data Fig. 2c,d). Next, we assessed the likelihood of recovering shared lineage barcodes from all pairs of cell classes and quantified lineage coupling by calculating a z-score for clone counts with respect to a random distribution (see Methods)²⁴. Hierarchical clustering of the pairwise correlation between coupling scores revealed structured groups, which comprised clonally related cell classes (Fig. 1e). STICR^{E10} and STICR^{E14} clones stayed predominantly within a class (Fig. 1e). However, STICR^{E10} showed stronger coupling between oligodendrocytes and oligodendrocyte precursor cells (OPCs), whereas STICR^{E14} showed stronger coupling between oligodendrocytes, OPCs and astrocytes (Fig. 1e). Notably, 37.2% of STICR^{E10} clones contained cells of a single class, compared with 62% of STICR^{E14}-descent clones (Fig. 1f). Moreover, 57.2% of STICR^{E10} clones were glial clones, 19% were neuronal clones and 23.8% were mixed clones (that is, clones with cells spanning multiple cell classes) versus 21.7%, 62% and 16.3%, respectively, for STICR^{E14} (Fig. 1g). STICR^{E10} mitotic progenitors produced larger clones, with up to 30 sister cells per clone, compared to STICR^{E14} mitotic progenitors, which produced up to 13 sister cells (Fig. 1h). Together, early-labelled progenitors generated a higher proportion of clones that dispersed across multiple cell classes than late-labelled progenitors that produced a majority of neuronal clones, consistent with progressive temporal fate specification of progenitors.

Clonal convergence and divergence

Our analysis thus far focused on the lineage relationships among cell classes. To gain higher resolution, we next explored clonal relationships between finer-grained subtypes. The 41 clusters were annotated on the basis of marker gene expression and

mapped to anatomical brain regions using Visium Spatial Gene Expression³³ (Extended Data Figs. 3, 4, Supplementary Data 3). Of these, nine clusters were reclustered to gain a higher level of detail (for example, cluster 7 was split into clusters 7a and 7b; Fig. 2a). Hierarchical clustering of the pairwise correlation between coupling z-scores revealed structured groups of clusters ('clonal groups' a–y; Extended Data Fig. 5).

Fig. 2: Lineage convergence and divergence in the mouse forebrain.

 **figure 2**

a, UMAP plot of single cells from the forebrain coloured by cluster. **b**, Schematics of patterns of lineage convergence and divergence. **c**, Dendrograms representing cluster relationships based on transcriptomic similarity (left) and lineage coupling correlations (STICR^{E14}, right). In purple are examples of lineage convergence; in green are

examples of lineage divergence. CGE, caudal ganglionic eminence; EAC, central extended amygdala; IN, interneuron; MGE, medial ganglionic eminence; NP, neuronal precursor; PN, projection neuron; VIP, vasoactive intestinal polypeptide. **d**, UpSet plot of selected intersections for STICR^{E14}. Only dispersing clones are shown. The bar graph at the top shows the number of observed intersections, and the bar graph on the right shows the number of cells per cluster. **e**, Schematic for lineage convergence of astrocyte clusters 3 and 12a. **f**, UpSet plot for selected ventral inhibitory neuron types in STICR^{E14}. The bar graph at the top shows the number of observed intersections, and the bar graph on the right shows the number of cells per cluster.

First, we tested whether excitatory and inhibitory neurons in the neocortex originate from the same progenitor. We identified clusters of cortical excitatory and inhibitory neurons based on the expression of canonical marker genes (*Gad1* inhibitory and *Slc17a7* excitatory) and Visium Spatial Gene Expression, and quantified their clonal relationship (Extended Data Fig. 6). In contrast to Delgado et al.²⁶ in the same issue, who found clonally related GABAergic and glutamatergic neurons in humans, we observed no evidence for shared lineages of excitatory and inhibitory cortical neurons in STICR^{E10} or STICR^{E14} mice (Extended Data Fig. 6c).

Next, we tested whether developmental histories can be predicted from the assumption that cell types with transcriptomically similar identities are clonally related. This assumption may not always hold true, because similar cell states could arise from different lineages (that is, convergence) and distinct cell states could share a lineage (that is, divergence) (Fig. 2b). To explore these possibilities, we compared hierarchies constructed from transcriptome information and hierarchies constructed from lineage coupling correlations (Fig. 2c). Seven astrocyte clusters appeared transcriptomically similar to one another, as they occupied a single clade on the transcriptome dendrogram, but largely belonged to different clades on the lineage dendrogram (Fig. 2c, clusters highlighted in purple). Moreover, we found that astrocyte lineages of the dorsal forebrain and astrocyte lineages of the ventral forebrain converged on seemingly identical astrocyte subtypes. In particular, clusters ‘3 Astrocyte *Slc12*’ and ‘12a Astrocyte *Ntrk2*’ were both clonally related to clusters ‘6 Inhibitory neuron OB *Synpr*’ and ‘5a Excitatory neuron upper cortex’ (Fig. 2d, Extended Data Fig. 7), which suggests that distinct ventral and dorsal radial glia can give rise to transcriptomically similar astrocyte populations (Fig. 2e). Another example of convergence was OB neuroblasts (‘4 Inhibitory NP OB *Ccnd2*’) and dentate gyrus neuroblasts (‘17 Excitatory NP *Sox11*’), which occupied a single clade on the transcriptome dendrogram but distant clades on the lineage dendrogram (Fig. 2c). Thus, cells originating from distinct mitotic progenitors located in different brain areas can converge to similar transcriptomic identities.

The most striking example of clonal divergence was observed for inhibitory neuron clones, which are known to derive from mitotic progenitors in the ganglionic eminences of the ventral forebrain (Fig. 2c, clusters highlighted in green). In particular, we identified six GABAergic projection neuron and interneuron clusters of the subpallium and OB that displayed high lineage coupling (Fig. 2c, Extended Data Fig. 5, see ‘clonal groups’ h,u,v). Neurons within these clusters included direct (D1) and indirect (D2) spiny projection neurons (SPNs) of the striatum (clusters 7b and 7a, respectively), projection neurons of the central extended amygdala (cluster 34), intercalated cells (ITCs) of the amygdala (cluster 8) and OB interneurons (clusters 2 and 6; Extended Data Fig. 8a,b). These GABAergic neurons were clonally related, although they showed drastically different transcriptomic profiles (Fig. 2c,f, Extended Data Figs. 7, 8a) and are known to have different morphologies, connectivity patterns and occupy different brain regions³. For example, ITCs of the amygdala were clonally related to interneurons of the OB and multiple GABAergic projection neuron types, including SPNs of the striatum and central extended amygdala. Moreover, both D1 and D2 SPNs were clonally related to interneurons of the OB (Fig. 2f). Next, we created three additional STICR^{E12} datasets of anatomically dissected brain regions (the OB, striatum and amygdala) and measured the similarity between the average gene expression of clusters (Extended Data Fig. 8c,d). Clones of GABAergic neurons were distributed across different forebrain structures and frequently across transcriptomically diverse subtypes. Thus, individual progenitors of GABAergic neurons can give rise to a wide range of different GABAergic subtypes (Extended Data Fig. 8e).

Embryonic divergence of GABAergic neurons

To test whether clonal divergence of GABAergic subtypes is the result of early fate specification within embryonic progenitor zones, or rather emerges during postnatal development, we studied single-cell lineage histories when molecular diversity of cell types first occurs. Because STICR labels mitotic progenitors indiscriminately along the embryonic ventricles, it is not suited to deliver a large number of lineage tags to a spatially defined region. To tag mitotic progenitors specifically in the ganglionic eminences, we developed a transposon-based barcoding approach (TrackerSeq; Fig. 3a, Extended Data Fig. 9a-d; see [Methods](#)) that uses the *piggyBac* transposon system to randomly integrate an eGFP reporter cassette into the genome of electroporated mitotic progenitors^{24,34}. A DNA sequence containing random nucleotides was cloned into the 3' untranslated region of eGFP, making it detectable by scRNA-seq.

Fig. 3: Newly born GABAergic sister cells diverge into different precursor states.

 **figure 3**

a, Schematic of the TrackerSeq experimental workflow. PBase, *piggyBac* transposase. **b**, Images of coronal brain sections electroporated with TrackerSeq^{E12} and collected at E14.5. Cx, cortex; GE, ganglionic eminence. Magnification on the bottom right panel shows a radial cluster of newborn cells (white arrowheads). **c**, UMAP plot of integrated embryonic scRNA-seq datasets, coloured by clusters. i, inhibitory; m, mitotic. **d**, Heatmap showing the normalized expression of the top ten marker genes for the five precursor states. **e**, Schematic of the strategy for computationally mapping embryonic precursor state cells to postnatal clusters. **f**, UMAP of the embryonic dataset, with precursor state cells coloured based on the mapping results. **g**, Bar graph quantifying the correlation-based mapping of cells from the five precursor states to selected postnatal ventral GABAergic neuron clusters. The numbers on the bars indicate the dominant mapped postnatal cluster. Inhib., inhibitory; VS, ventral striatum. **h**, UpSet plot for all intersections of TrackerSeq^{E12}. The bar graph on the top shows the number of intersections. Mitotic clusters were merged in a single cluster (mitotic). The total cell number per cluster is represented in the bar graph on the right. Intersections among precursor states are coloured in ochre. The bars are colour-coded according to Fig. 3c. The inset in the top right represents the percentage of multicellular clones that follow a single trajectory or dispersed across several

precursor state trajectories. **i**, Examples of clones where sibling cells traverse a single developmental trajectory (left) or different trajectories (right) on the UMAP.

We targeted TrackerSeq to ganglionic eminence progenitors at E12.5, FACS-enriched electroporated cells at E16.5, and performed scRNA-seq (Fig. [3a, b](#)). Hierarchical clustering of TrackerSeq DNA tags organized cells into 256 distinct multicellular clones of GABAergic neurons (Extended Data Fig. [9e–h](#)). To gain a high resolution of embryonic cell states in the ganglionic eminences, we integrated the TrackerSeq datasets with wild-type scRNA-seq datasets that we collected at E13.5 and E15.5 from the medial, caudal and lateral ganglionic eminences (Fig. [3c](#), Extended Data Fig. [9e, f](#)). We performed a pseudotime trajectory analysis using Monocle3^{[35](#)}, which is a diffusion pseudotime algorithm that learns the sequence of gene expression changes and identifies developmental branch points (Extended Data Fig. [9i](#)). In addition, we used RNA velocity^{[36](#)}, a method that compares the ratios of unspliced and spliced mRNA per gene, to predict the direction and speed of cell-state transitions across the trajectories (Extended Data Fig. [9j](#)).

From a common pool of mitotic progenitors, five different trajectories (that is, precursor states) of postmitotic inhibitory neurons emerged soon after cell-cycle exit, which we named after the top marker genes that these clusters expressed (*'i_Six3/GucylA3'*, *'i_Ebf1/Isl1'*, *'i_Phlda1/Isl1'*, *'i_Nr2f2'* and *'i_Nxphl'*; Fig. [3c, d](#)). We used a correlation-based distance metric (see [Methods](#)) to map cells from each embryonic precursor state to inhibitory clusters of the postnatal STICR dataset (Fig. [3e–g](#), Extended Data Fig. [9k](#)). For example, 83% of cells from the trajectory *'i_Six3/GucylA3'* mapped to the postnatal cluster '7a D2 SPNs', and 89% of cells from trajectory *'i_Ebf1/Isl1'* mapped to cluster '7b D1 SPNs' (Fig. [3g](#)), suggesting an early emergence of postnatal signatures. Consistent with this idea, OB interneuron precursors, as well as D1 and D2 striatal precursors, maintained multiple marker genes through development (Extended Data Fig. [9l](#)).

We next asked whether clonally related cells traverse the same or different trajectories. Notably, while cells of 63.6% of clones entered the same trajectory, 36.4% of the clones diverged into differing trajectories shortly after they exited the cell cycle (Fig. [3h, i](#)). For example, we found sister cells located on the '7b D1 SPN' and '7a D2 SPN' trajectories, '7a D2 SPN' and '8 Inhibitory ITC–amygdala' trajectories, and the '2 Inhibitory neuron OB *Meis2*' and '8 Inhibitory ITC–amygdala' trajectories (Fig. [3h, i](#)). Taken together, these data show that progenitor cells in the ganglionic eminences can produce daughter cells that traverse different developmental trajectories during peak neurogenesis. This suggests that clonal divergence into different GABAergic precursor states is initiated at the level of mitotic progenitor cells and thus as a lineage-dependent process (Extended Data Fig. [10](#)).

Discussion

During neurogenesis, a wide range of postmitotic neurons and glia arise from mitotic progenitors lining the embryonic ventricles. The extent to which developmental histories of mature cell types can be predicted on the basis of the assumption that cell types with transcriptomically similar identities are developmentally related has remained obscure. Using methods that simultaneously capture transcriptomic signatures and lineage histories of individual cells, we found both transcriptomically similar cell types that arose from different lineages (that is, convergence) and transcriptomically distinct cell types that share a clonal relationship (that is, divergence). The most striking example of divergence was GABAergic clones stemming from mitotic progenitors in the ventral forebrain. These clones did not only disperse into distinct brain regions, but also diverged into types with grossly different transcriptomic signatures. While perhaps some degree of clonal divergence would be expected early in neurogenesis at a time when a large number of radial glial progenitors undergo symmetric proliferative divisions, our results reveal clonal divergence at later stages of neurogenesis.

In agreement with Mayer et al.¹, we found that in all ganglionic eminences, newborn GABAergic neurons diverge into different precursor states after cell-cycle exit. It was not clear whether clonally related sister cells enter the same or different precursor states (Extended Data Fig. 10). Because our new methods capture clonal histories, we were able to address this question and found that newborn sister cells often diverged in different trajectories, suggesting that mechanisms on the level of progenitors delineate genetic identity and ultimately cell fate^{37,38}. The ganglionic eminences can be divided into more than a dozen progenitor subdomains that are uniquely demarcated by the combinatorial expression of transcription factors^{39,40}. The superimposition of a cellular mechanism that gives rise to different postmitotic signatures with domain-specific factors and morphogens provides a large combinatorial framework that could explain how an enormous diversity of inhibitory types is generated in the forebrain. Whether the sequential production of different types follows a stereotypic sequence, or involves stochastic events that occur during cell-cycle exit, remains unknown. One possibility is that the sequential production of cell types depends on the interaction of progenitors with developmentally dynamic transcription factors and morphogens.

While recent work has examined how cell populations vary across species⁴¹, the methods developed in our study and Delgado et al.²⁶ in the same issue enable comparison of developmental histories of cell types across mouse and human. The identification of species-specific lineages will provide insight into the evolution of cellular diversity. Both STICR and TrackerSeq capture partial clones, which is sufficient to detect clonal divergence and convergence. However, at present, our methods can infer but not prove lineage restriction.

Somatic mutations, which are clonally inherited, contribute to numerous neurodevelopmental diseases^{42,43,44,45,46,47}. Determining lineage relationships might explain why certain subsets of cells are affected in developmental disorders. Looking forward, we anticipate that studies combining lineage information with genetic mutations will facilitate the explorations of such clinical phenotypes.

Methods

STICR lentiviral library preparation and validation

We synthesized a high-complexity lentivirus barcode library that encodes approximately 60–70 million distinct oligonucleotide RNA sequences (STICR barcodes). STICR barcodes comprised three distinct oligonucleotide fragments cloned sequentially into a multicloning site within the 3' UTR of an enhanced green fluorescent protein (eGFP) transgene under control of a ubiquitous CAG promoter in a modified lentiviral plasmid (pSico, Addgene, #11578). Each barcode fragment was derived from one of three oligonucleotide pools comprising 500 distinct sequences, allowing for up to 125 million unique combinatorial barcode sequences (500^3). Following the ligation of each oligonucleotide fragment into the multicloning site, the plasmid library was electroporated into electrocompetent MegaX DH10B cells (Fisher, #C640003) and grown for 12 h overnight at 37 °C on LB agar plates (Fisher, BP1425-500) with carbenicillin (Fisher, #BP26481). The resulting colonies were scraped and plasmid extraction was performed using a midi-prep kit (Macherey Nagel, 740412.5). This process was repeated until all three barcode fragments were added.

Lentivirus production was performed by first transfecting HEK293 cells with the barcode library along with lentiviral helper plasmids pMDLg/pRRE (Addgene, #12253), pRSV-Rev (Addgene, #12253) and envelope protein VSV-G (Addgene, #8454) using JetPrime (PolyPlus, 114-07). HEK293 media were changed 24 h after transfection and replaced with 35 ml Ultraculture media (Lonza BE12-725F), 350 µl sodium pyruvate (11 mg/ml stock, Thermo Fisher, 11360070), 350 µl sodium butyrate (0.5 M stock, Sigma, B5887), and 350 µl antibiotic/antimycotic (Thermo, 15-240-062) (<http://syntheticneurobiology.org/protocols/protocoldetail/31/12>). After an additional 48 h, media were collected, concentrated with an ultracentrifuge, and then resuspended in 50–100 µl of sterile PBS.

To confirm that transcribed STICR barcodes can be accurately recovered using scRNA-seq, we performed a ‘barnyard experiment’ in which we infected separate cultures of human cortical cells (GW18 sample) and mouse 3T3 cells (ATCC) with different STICR libraries. These libraries could be distinguished from each other by a constant sequence unique to each library (‘viral index’). After 3 days, we dissociated cultures with papain and FACS-isolated eGFP⁺ cells. eGFP⁺ cells from both species

were then mixed together and loaded into a 10X Genomics Chromium Single Cell '3 prime kit (10x Genomics, PN-100007). Following sequencing, transcript libraries were aligned with CellRanger (version 3.0.2) to a hybrid mouse/human genome and droplets were determined to be either a mouse cell, human cell or multiplet. STICR barcodes were recovered (see below) and the recovered viral index sequence was used to match recovered barcode to the barcode initially used to infect each experiment. Finally, we quantified recovered viral indices for mouse, human and multiplet droplets.

To measure STICR plasmid library barcode diversity, we first digested 1 µg of each library with XhoI and then ligated a PCR adapter containing a unique molecular identifier (UMI) to this site. Ligation products were amplified by PCR using Q5 Hot Start High Fidelity 2x Master Mix (NEB, #M0494) using primers targeting the STICR sequencing primer site and the adapter sequence using the following program: (1) 98 °C for 30 s, (2) 98 °C for 10 s, (3) 62 °C for 20 s, (4) 72 °C for 10 s, (5) repeat steps 2–4 15 times, (6) 72 °C for 2 min, and (7) 4 °C hold. Following PCR amplification, a 0.8–0.6× dual-sided size selection was performed using Ampure XP beads (Beckman Coulter, #A63881). The resulting libraries were sequenced to the depth of approximately 30 million reads. STICR barcode sequences were extracted using custom scripts that removed PCR duplicate reads using the UMI (see below in 'ScRNA-seq analysis and STICR barcode analysis' for a general description). Since it is prohibitively expensive to sequence the library to saturation, we extrapolated the total number of unique STICR barcodes using the Preseq⁴⁸ command lc_extrap and default settings. Together with the measured relative barcode abundances, we used the extrapolated STICR barcode library size to model barcode collisions using the R (v4.0.1) programming language. Using base R functions, we simulated the labelling of a starting population of cells with a range of sizes from 10^1 to 10^6 and repeated each simulation 20,000 times. We then quantified the mean number of unique barcodes chosen for each starting cell population size. The difference between the starting cell population size and the number of unique barcodes present represented the number of collisions that had happened at that population size.

TrackerSeq library preparation and validation

TrackerSeq is a *piggyBac* transposon-based³⁴ library, developed to be compatible with the 10x single-cell transcriptomic platform. It records the *in vivo* lineage history of single cells through the integration of multiple oligonucleotide sequences into the mouse genome. Each of these individual lineage barcodes is a 37-bp long synthetic nucleotide that consists of short random nucleotides bridged by fixed nucleotides. This design results in a library with a theoretical complexity of approximately 4.3 million lineage barcodes (16^8) with each barcode differing from another by at least 5 bp.

To construct the library, the *piggyBac* donor plasmid (Addgene #40973) was altered to include a number of modifications. A Read2 partial primer sequence was cloned into the 3' UTR of the eGFP to enable retrieval by the 10x platform. The sucrose gene was cloned into the vector, so that empty plasmids that fail to incorporate a lineage barcode during the cloning process are removed. Following digestion with BstXI to remove the sucrose gene, the plasmid was run on a gel and column purified. The lineage barcode oligo mix was cloned downstream of the Read2 partial primer sequence in the purified donor plasmid via multiple Gibson Assembly reactions, as previously described⁴⁹. Gibson assembly reactions (NEB, #E2611S) were then pooled and desalted with 0.025 µm MCE membrane (Millipore, #VSWP02500) for 40 min, and finally concentrated using a SpeedVac. 3 µl of the purified assembly is incubated with 50 µl of NEB10-β-competent *Escherichia coli* cells (NEB, #C3019H) for 30 min at 4 °C, then electroporated at 2.0 kV, 200 Ω, 25 µF (Bio-Rad, Gene Pulser Xcell Electroporation Systems). Electroporated *E. coli* were incubated for 90 min shaking at 37 °C and then plated into pre-warmed sucrose/ampicillin plates. The colonies were scraped off the plates after 8 h, and the plasmids were grown in LB medium with ampicillin up to OD = 0.5. The plasmid library was purified using column purification kit (Zymo Pure II Plasmid Maxiprep kit, #D4202). We first assessed the integrity of the TrackerSeq barcode libraries by sequencing the library to a depth of approximately 42 million reads to test whether any barcode was over-represented. Around 3.6 million valid lineage barcodes that had a quality score of 30 or higher were extracted from the R2 FASTQ files using Bartender⁵⁰. One thousand barcodes were randomly sampled from the extracted lineage barcodes to assess hamming distance. To group similar extracted barcodes into putative barcodes, Bartender assigns a UMI to each barcode read to handle PCR jackpotting errors, and clusters them. The cluster distance was set to 3 so that extracted barcodes within 3 bp of each other have a chance of being clustered together. A total of 2×10^5 clusters of barcodes were identified, suggesting that the barcode library has a diversity that is at least in the 10^5 range.

Mice and in utero surgeries

All mouse colonies were maintained in accordance with protocols approved by the Bavarian government at the Max Planck Institute of Neurobiology or the IACUC at the NYU Grossman School of Medicine. Swiss Webster and C57BL/6 wild-type females were used, and embryos were staged in days post-coitus, with E0.5 defined as 12:00 of the day a vaginal plug was detected after overnight mating. Timed pregnant mice were anaesthetized with isoflurane (5% induction, 2.5% during the surgery) and treated with the analgesic Metamizol (WDT). In utero surgery and injection of the STICR lentiviral library in the lateral ventricles of the embryonic mouse forebrains at E10.5–E14.5 were performed as previously described⁵¹. A microsyringe pump (Nanoject III Programmable Nanoliter Injector (100/240V) (#DRUM3-000-207)) was used to inject approximately 0.5 µl of the STICR library per embryo. For embryos

injected at E10.5, ultrasound backscatter microscope (UBM) was used to allow for image-guided injections. For in utero electroporation of the TrackerSeq library, E12.5 embryos were injected unilaterally with 700 nl of DNA plasmid solution made of 0.5 μ g/ μ l pEF1a-pBase (*piggyBac*-transposase; a gift from R. Platt) and the TrackerSeq library 0.5 μ g/ μ l, diluted in endo-free TE buffer and 0.002% Fast Green FCF (Sigma), into the lateral ventricle via a microsyringe pump. Embryos were then electroporated by holding each head between platinum-plated tweezer electrodes (5 mm in diameter, BTX, #45-0489) across the uterine wall, while 5 electric pulses (35 V, 50 ms at 1 Hz) were delivered with a square-wave electroporator (BTX, ECM 830)⁵². Pregnant dams were kept in single cages and pups were kept with their mothers, in the institutional animal facility under standard 12:12-h light–dark cycles, at a room temperature of 72° F \pm 2° F and a humidity of 30–70%.

Sample collection

Virally injected brains were collected from mouse pups between ages postnatal day 5 and day 15 (P5–P15) (Supplementary Data 1). Brains were dissected in ice-cold pre-bubbled artificial cerebrospinal fluid (aCSF), and sectioned into 400- μ m coronal sections on a Leica VT1200S Vibratome. Coronal brain sections were then dissected such that the forebrain was collected, thus excluding the thalamus, hypothalamus, brainstem and cerebellum. Alternatively, OBs, amygdalae and striata were manually dissected out from sliced brains, and processed separately. Collected tissue was then dissociated with the Miltenyi BioTech Neural Tissue Dissociation Kit (P) (#130-092-628) on the gentleMACS Dissociator according to the protocol of the manufacturer. To isolate and collect virally infected cells, flow cytometry was done using a SY3200 Cell Sorter (software WinList 3D version 8.0.) or BD FACSAria III Cell Sorter (BD FACSDiva Software, version 8.0.2) with a 100- μ m nozzle. The cell suspensions were first gated on forward scatter, then within this population based on eGFP expression. eGFP-expressing cells were collected in bulk for downstream processing on the 10x Genomics Chromium platform.

Ganglionic eminences were collected from mouse embryos at E13.5 and E15.5 (Supplementary Data 1) according to the following protocol: embryos were removed from the uterus of wild-type Swiss Webster females, and stored in ice-cold L-15 medium. Brains were removed from the embryonic skulls, and the MGE, CGE and LGE were dissected out. MGEs, CGEs and LGEs were then each pooled together from multiple embryos, so that each eminence type was processed independently, and dissociated with the Miltenyi BioTech Neural Tissue Dissociation Kit (P) (#130-092-628) on the gentleMACS Dissociator according to the protocol of the manufacturer.

For embryonic lineage tracing, we collected electroporated brains from mouse embryos at E16.5 (Supplementary Data 1) in Leibowitz medium with 5% FBS. Papain

dissociation system was carried out according to the recommended protocol (Worthington, #LK003150), and to isolate positive cells, flow cytometry was done using a BD FACSAria III Cell Sorter (BD FACSDiva Software, version 8.0.2) with a 100- μ m nozzle. For all FACS experiments, non-eGFP-expressing brain tissue was used as a negative control for excluding background fluorescence.

Immunohistochemistry

E14.5 and P10 mice were perfused with 4% PFA and post-fixed overnight in 4% PFA at 4 °C. Coronal sections (60 μ m) were performed using vibrating microtome. Immunofluorescent staining was performed as follows: sections were incubated for 1 h at room temperature in blocking solution (5% BSA and 0.3% Triton-X100 in PBS), then overnight at 4 °C with primary antibodies. Sections were rinsed three times in PBS 1X and incubated for 1 h at room temperature with the corresponding secondary antibody (1:500, Life Technologies). Three washes with PBS 1X were performed, the second wash using Hoechst staining solution (1:10,000 in PBS 1X, Invitrogen) to label nuclei, before dry mounting on slides with Fluoromount-G (Invitrogen). For imaging, the primary somatosensory area was used as region of study. Images were acquired on a Leica Sp8 confocal laser scanning microscope.

The primary antibodies used included: rabbit anti-CUX1 1:500 (Santa Cruz, #SC13024), rabbit anti-GABA 1:2,000 (Merck, #A2052), rabbit anti-GFP 1:1,000 (Invitrogen, #A11122), rabbit anti-Iba1 1:500 (Wako, #019-19741), rabbit anti-OLIG2 1:500 (Merck, #AB9610), rabbit anti-S100 β 1:500 (Merck, #S2644), and rat anti-CTIP2 [25B6] 1:500 (Abcam, #AB18465).

The secondary antibodies used were: 647 Alexa Fluor plus goat anti-rabbit (Invitrogen, #A32733), 555 Alexa Fluor goat anti-rat (Invitrogen, #A21434), and 555 Alexa Fluor goat anti-rabbit (Invitrogen, #A21428).

Preparation of RNA-seq, STICR and TrackerSeq libraries

For experiments utilizing the 10x Genomics platform, the following reagents were used: Chromium Single Cell 3' Library & Gel Bead Kit v2 (PN-120237), Chromium Single Cell 3' Chip Kit v2 (PN-120236) and Chromium i7 Multiplex Kit (PN-120262) were used according to the manufacturer's instructions in the Chromium Single Cell 3' Reagents Kits V2 User Guide; Chromium Single Cell 3' Library & Gel Bead Kit v3 (PN-1000075), Chromium Single Cell 3' Chip Kit V3 (PN-1000073) and Chromium i7 Multiplex Kit (PN-120262) were used according to the manufacturer's instructions in the Chromium Single Cell 3' Reagents Kits V3 User Guide; Chromium Single Cell 3' Library & Gel Bead Kit v3.1 (PN-1000268), Chromium Single Cell 3' Chip Kit V3.1 (PN-1000127) and Dual Index Kit TT Set A (PN-1000215) were used according to the

manufacturer's instructions in the Chromium Single Cell 3' Reagents Kits V3.1 User Guide (Dual Index).

The lineage barcode library retrieved from RNA was amplified with a standard NEB protocol for Q5 Hot Start High-Fidelity 2X Master Mix (#M094S) in a 50- μ l reaction, using 10 μ l of cDNA as template. Specifically, each PCR contained: 25 μ l Q5 High-fidelity 2X Master Mix, 2.5 μ l 10 μ M P7_indexed reverse primer, 2.5 μ l 10 μ M i5_indexed forward rimer, 10 μ l molecular grade H₂O, 10 μ l cDNA (for primer sequences and indices, see Supplementary Data 1). The PCR protocol for amplifying STICR lineage libraries was: (1) 98 °C for 30 s, (2) 98 °C for 10 s, (3) 62 °C for 20 s, (4) 72 °C for 10 s, (5) repeat steps 2–4 11–18 times, (6) 72 °C for 2 min, and (7) 4 °C hold. The PCR protocol for amplifying TrackerSeq lineage libraries was: (1) 98 °C for 30 s, (2) 98 °C for 10 s, (3) 63 °C for 20 s, (4) 72 °C for 10 s, (5) repeat steps 2–4 11–18 times, (6) 72 °C for 2 min, and (7) 4 °C hold. Libraries were purified with a dual-sided SPRI selection using Beckman Coulter Agencourt RNAClean XP (A63987), and quantified with an Agilent BioAnalyzer. Some STICR libraries (DI_T_199, DI_T_203, DI_T_211, DI_T_222, DI_T_233, DI_T_238, DI_T_239, DI_T_240, DI_T_241, DI_T_242, DI_T_287, DI_T_289, DI_T_304 and DI_T_305) were constructed and sequenced twice to achieve higher resolution.

Sequencing and read mapping

Transcriptome and barcode libraries were sequenced either on an Illumina NextSeq 500 at the Next Generation Sequencing Facility of the Max Planck Institute of Biochemistry, at the Genomics Core Facility at the Helmholtz Center in Munich, or on a NovaSeq at the Broad Institute. For a detailed report on each dataset, see Supplementary Data 1. Sequencing reads in FASTQ files were aligned to a reference transcriptome (mm10-2.1.0) and collapsed into UMI counts using the 10x Genomics Cell Ranger software (version 3.0.2 or 5.0.1).

Processing of STICR barcode reads

STICR barcode analysis was performed using custom scripts. First, BBMap (BBMap —Bushnell B.; sourceforge.net/projects/bbmap/) was used to remove low-quality reads and then extract reads containing STICR barcode sequences. Then, BBMap was used to extract individual STICR barcode fragments, which were then aligned to our pre-defined fragment reference sets using Bowtie (v5.2.1)⁵³, allowing for up to two mismatches per fragment. Aligned STICR barcodes were compiled into a file containing their corresponding 10X cell barcode and 10X UMI sequences using Awk. Finally, UMI-tools (v.0.5.1)⁵⁴ was used to remove duplicate STICR barcode or cell barcode reads by UMI, allowing for 1 bp mismatch in the UMI. STICR barcodes—CBC pairings with at least five distinct UMIs were retained for clonal analysis.

Among cells with multiple STICR barcodes passing these criteria, we attempted to find a ‘dominant’ STICR barcode that we defined as containing greater than or equal to five times the number of UMI counts than the next most abundance STICR barcode. Dominant STICR barcodes meeting these criteria were considered to be the clonal barcode of their respective cells and retained for further analysis. Within metadata files, these CBC–STICR barcode pairings are referred to as tier 2, whereas CBC–STICR barcode pairs with only a single STICR barcode meeting threshold criteria are referred to as tier 1. Only CBCs associated with a single STICR barcode tier 1) were used in this study, with the exception of the OB–amygdala–striatum experiment in Extended Data Fig. 8, where tier 2 was used. Finally, we applied an additional UMI threshold to STICR barcodes, requiring that all STICR barcode–CBC pairings used for clonal analysis have at least nine distinct UMIs. Three STICR barcodes (Index1_Bit1_F_083-Bit2_F_060-Bit3_F_055, _Bit1_F_057-Bit2_F_103-Bit3_F_244 and IndexE_Bit1_F_246-Bit2_F_178-Bit3_F_497) were removed from downstream analyses, because they were present in more than one dataset. Two of the total 21 STICR datasets (CA199 and CA233) contained only single-cell clones. These datasets were retained to support single-cell cluster analysis and clone size quantification.

Processing of TrackerSeq barcode reads

Reads in the R2 FASTQ files were pre-processed so that the sequences to the left and right of the lineage barcodes (BC) were trimmed. Lineage barcodes shorter than 37 bp were discarded. Cell barcodes (Cell) were extracted from the corresponding Seurat object of the dataset to generate a cell barcode whitelist. The extracted cell barcodes and UMIs were added to the read names of the lineage barcode FASTQ files. The resulting FASTQ files were processed to output a sparse matrix in csv format, where rows were cells identified by individual cell barcodes and columns were lineage barcodes. Only Cell–UMI–BC triples supported by at least 10 reads and Cell–BC pairs with at least 6 UMI were considered for further analyses. CloneIDs were assigned to cell barcodes by clustering the matrix using Jaccard similarity and average linkage as demonstrated by Wagner and colleagues²⁴. The resulting dendrogram was cut at a height of 0.999 to obtain the clonal groupings. The clonal groupings showed that there were 4,282 barcodes distributed over 2,370 cells in the total dataset, where 56.0% of them were marked by 2 or more barcode integrations, and 8.4% of them were marked by 5 or more integrations in the total dataset. Among the inhibitory neurons featured in Fig. 3, these numbers were 85.7% and 9.5%, respectively.

Cell filtering, data normalization, batch correction and clustering of STICR datasets

The Seurat workflow (version 3.1.4) was used for cell filtering, data normalization and cluster identification in scRNA-seq datasets. Data were read into R (version 3.6.0) as a

count matrix. Each dataset was filtered with cut-offs for: maximum and minimum gene expression, maximum nCount_RNA, and the percentage of total reads that aligned to the mitochondrial genome (for applied cut-offs, see Supplementary Data 1). Filtered data were then used for standard processing with Seurat. Unless otherwise indicated, gene expression values for each cell were divided by the total number of transcripts and multiplied by 10,000. These values were then log transformed using log1p via the NormalizeData() function. Genes were scaled and centred using the ScaleData() function. We used Harmony (v1.0)²⁸ within the Seurat workflow with default parameters (theta = 2, lambda = 1, sigma = 0.1) to integrate different STICR datasets. We used the first 35 Harmony embeddings for UMAP (<https://github.com/lmcinnes/umap>) visualizations and clustering analysis.

To partition cells into clusters, we constructed a shared-nearest neighbour graph based on Harmony embeddings via the FindNeighbors() function to use as input to the SLM algorithm, implemented through the FindClusters() function in Seurat (dimensions = 35, res = 1). Cluster-specific marker genes were identified by comparing cells of each cluster to cells from all other clusters. Genes were considered differentially expressed based on fold change, minimum expression and adjusted *P* value cut-offs (Supplementary Data 3). The Wilcoxon rank sum test was implemented via the Seurat function FindAllMarkers().

Clusters were manually annotated based on marker gene expression, spatial transcriptome mapping, as well as publicly available databases, primarily DropViz (dropviz.org)⁵⁵ and Mouse Brain Atlas (<http://mousebrain.org/genearch.html>)⁵⁶. Of the 41 unsupervised STICR clusters, 9 clusters were reclustered to gain a higher level of detail (for example, cluster 7 was split into clusters 7a and 7b; Fig. 3a). More specifically, clusters were isolated using subset() and clustered again using FindClusters(). Sub-clusters that could not be assigned to a cell type were assigned ‘unknown’ and excluded from the lineage analysis (0.76% of the total cells). Cluster 22, which contained sub-clusters of different classes, was labelled ‘mixed’. All analyses were carried out based on refined clusters, except for panels illustrating the spatial analysis.

Lineage analysis of cell classes in STICR

To quantify clonal relationships between cell classes, Seurat clusters were merged into cell classes (Fig. 1d) based on the co-expression of multiple marker genes (neurons (*Tubb3* and *Mef2c*); neuronal precursors (*Gad1* and *Neurod2*); mitotic cells (*Ube2c* and *Top2a*); astrocytes (*Aldh1l1* and *Gfap*); oligodendrocytes (*Olig1* and *Plp1*); OPCs (*Pdgfra* and *C1ql1*); vascular cells (*Rgs5*); epithelial cells (*Ttr*); ependymal cells (*Tmem212*); and macrophages (*Ccl4* and *C1qa*)) (Extended Data Fig. 2a, Supplementary Data 2). Clones were categorized as containing sister cells that were

glia only (astrocyte, OPC and oligodendrocyte classes), neuron only, or glia and neuron mix (neuron, astrocyte, OPC and oligodendrocyte classes), and the number of clones in each of these three categories was quantified relative to the total number of clones at each developmental stage.

Lineage coupling z-scores and correlations forebrain-wide STICR datasets

The numbers of shared clones, as well as lineage coupling z-scores and correlations were calculated for each pair of cell states based on the methods outlined by Wagner and colleagues²⁴ as follows:

1. 1.

Definitions:

For a given clone c_1 with n cells, and a given cell-state pair $\{s_1, s_2\}$:

1. 1.1.

Let k be the number of cells of clone c_1 that were assigned to either of the cell states of the pair, that is, s_1 or s_2 .

2. 1.2.

Clone c_1 is defined to be ‘shared’ between states s_1 and s_2 if $k \geq 2$, and there was at least one cell of clone c_1 assigned to each state.

3. 1.3.

Let p be the fraction of clone c_1 that k represents, that is, $p = k / n$.

4. 1.4.

A metric for the cell-state pair $\{s_1, s_2\}$ is defined as the sum, over all ‘shared’ clones, of the p of each clone.

2. 2.

The metric defined in (1.4) was computed for each pair of cell states, according to the observed data.

3. 3.

A distribution of values of the metric was computed for each pair of cell states in the following way:

1. 3.1.

For $N = 10,000$ iterations, the following simulation was done:

1. 3.1.1.

Maintaining the observed distribution of the number of cells per cell state, the state assignments of the individual cells were randomly shuffled.

2. 3.1.2.

The metric defined in (1.4) was computed for each pair of cell states, according to the data resulting from this simulation.

4. 4.

For each pair of cell states, its lineage coupling z-score is defined as the z-score of its observed metric computed in (2), with respect to the distribution computed in (3).

Positive z-scores indicate pairs of cell states that shared more lineage barcodes than expected by chance, whereas a negative score indicates that a state pair was significantly less coupled than expected by chance.

5. 5.

For each pair of cell states, its lineage coupling correlation is defined as the correlation between all the lineage coupling z-scores of each individual cell state of the pair.

Dendrograms and UpSet plots

Dendrograms representing transcriptomic relationships were generated with the BuildClusterTree() function in Seurat, which constructs a phylogenetic tree relating the ‘average’ cell from each identity cluster. The tree is estimated on the basis of a distance matrix constructed in the gene expression space. Dendrograms representing lineage relationships were generated using the hclust() and dist() functions on lineage coupling correlations, with an average linkage clustering and Euclidean distance

metric. The interrelation between cell types can only be coarsely represented in hierarchical dendrograms. Dendrograms represent overall transcriptomic similarities and dissimilarities, but they fail to capture less obvious similarities between otherwise distinct cell types. Similarly, dendrograms may represent the general nexus of clonal relationships but overlook infrequent relationships.

UpSet plots were created in R using the UpSetR library⁵⁷. For set size, we used the number of cells per cluster.

Correlation-based distance measure for amygdala, OB and striatum datasets

Amygdala, OB and striatum datasets were pre-processed as mentioned above. Cell types were manually annotated and neuronal types were divided via subset(). The distance between the log-normalized average cluster gene expression was calculated using the Spearman correlation-based distance measure in the get_dist() function and visualised using fviz_dist() from the R package factoextra v1.0.7.

Cell filtering, data normalization batch correction and clustering of embryonic datasets

The Seurat pipeline (version 3.1.4) was used for cluster identification in scRNA-seq datasets. Embryonic transcriptome datasets (MUC28072, CA303, CA300, CA302, CA299, CA301 and CA298) were read into R (version 3.6.0) as a count matrix. Each dataset was filtered with cut-offs for: maximum or minimum gene expression, maximum nCount_RNA and the percentage of total reads that aligned to the mitochondrial genome (for applied cut-offs, see Supplementary Data 1). In addition, embryonic datasets were filtered with DoubletFinder version 2.0.3 (ref.⁵⁸).

We used regularized negative binomial regression⁵⁹ to normalize UMI count data for all embryonic datasets. Cells with UMI counts for Neurod2 > 2 and Neurod6 > 2, which are markers of excitatory neurons, were removed. The TrackerSeq dataset was clustered using Seurat standard procedures and clusters expressing marker genes for excitatory neurons were removed. We created an ‘integrated’ data assay including all embryonic datasets for downstream analysis as described by Stuart and colleagues⁶⁰. Clusters of cells were identified by a shared nearest neighbour modularity optimization-based clustering algorithm. Uniform manifold approximation and projection (UMAP) dimensional reduction (<https://github.com/lmcinnes/umap>) was applied to the integrated data assay for visualization.

Trajectory analysis of embryonic datasets

Trajectory inference and pseudotime calculations were done with Monocle3 (ref.³⁵). RNA velocity was estimated using the R library velocyto.R³⁶. 10x output files were preprocessed with velocyto, version 0.17.17 (<https://velocyto.org>) using the command velocyto run10x. The velocyto.R-package ‘velocyto.R’ version 0.6 was used for RNA velocity estimation in R.

Mapping embryonic cells to postnatal clusters

To map cells from embryonic trajectories to postnatal cell types, we first selected the five embryonic Seurat clusters from the ‘integrated’ data assay that were located at the tip of the Monocle trajectories, as well as Seurat clusters from the postnatal STICR dataset that were identified and annotated as subpallial GABAergic neuron types. We focused on 1,855 genes that were identified as variable features at both developmental stages using the Seurat FindVariableFeatures() function. For these genes, we averaged the log-normalized expression in the postnatal clusters to create postnatal cell-type model vectors. We then calculated Pearson correlations between all individual cells of the embryonic clusters and the model vectors as described in Mayer et al¹. We assigned each cell to the postnatal cluster with the highest correlation, but also calculated empirical *P* values to determine the significance of the assignment by permuting the single-cell data for a random background. We left the model vectors unchanged, but permuted the single-cell expression data 100 times. For each permutation and each cell, we kept track of the largest Pearson correlation to the model vectors, and calculated a *P* value for the cluster assignment by counting what fraction of correlation scores was larger than the one used for the cluster assignment. In a final step, we turned all *P* values into false discovery rates (FDRs) and mapped only cells with an FDR < 0.1 to the postnatal clusters.

Spatial gene expression in STICR

To infer the spatial location of the clusters, the STICR datasets were integrated with the Visium Spatial Transcriptomic datasets for sagittal and coronal sections of the mouse brain provided by 10x genomics (<https://support.10xgenomics.com/spatial-gene-expression/datasets>). We applied an ‘anchor’-based integration workflow in Seurat v3, which enables the probabilistic transfer of annotations from a reference to a query set. The spatial reference dataset and the lineage dataset were normalized using the SCTransform() function, which builds regularized negative binomial models of gene expression, and performed dimensionality reduction using the RunPCA() function and then performed label transfer using the functions FindTransferAnchors() and TransferData(). This procedure outputs, for each spatial spot, a probabilistic classification for each of the scRNA-seq-derived cell states. We added these predictions as a new assay in the Seurat object for visualization using the function SpatialFeaturePlot().

Scatter plots

The top 100 marker genes were calculated using the Seurat function FindMarkers () for a selection of GABAergic clusters in the postnatal STICR dataset and the merged embryonic dataset, respectively (postnatal: ‘2 Inhibitory neuron OB *Meis2*’, ‘6 Inhibitory neuron OB *Synpr*’, ‘7a D2 SPNs’, ‘7b D1 SPNs’, ‘8 Inhibitory ITC amygdala’, ‘34 Inhibitory PN ventral striatum/central extended amygdala (EAC)’; ‘13a MGE IN *Snhg11*’, ‘19a CGE VIP IN’, ‘19b CGE neurogliaform IN’, ‘13a MGE IN *Snhg11*’; embryonic: ‘i_ *Six3/Gucy1a3*’, ‘i_ *Ebf1/Isl1*’, ‘i_ *Phldal/Isl1*’, ‘i_ *Nr2f2*’, ‘i_ *Nxph1*’). On the basis of the correlation-based mapping of embryonic cells to postnatal clusters (see previous paragraph), we selected pairs of embryonic and adult clusters for the scatter plot. We plotted the SCT normalized average cluster gene expression of the top 100 marker genes from each stage.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The sequencing datasets generated for the current study are available in the Gene Expression Omnibus (GEO) under the accession number [GSE188528](#). Publicly available gene expression data used for cluster annotation can be accessed as follows: DropViz ([dropviz.org](#)) and Mouse Brain Atlas ([http://mousebrain.org/genesearch.html](#)). Visium Spatial Transcriptomic Datasets for sagittal and coronal sections of the mouse brain provided by 10x genomics ([https://support.10xgenomics.com/spatial-gene-expression/datasets](#)).

Code availability

The analyses described here are available on GitHub: [https://github.com/mayer-lab/Bandler-et-al_lineage](#).

References

1. 1.

Mayer, C. et al. Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).

2. 2.

Di Bella, D. J. et al. Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* **595**, 554–559 (2021).

3. 3.

Lim, L., Mi, D., Llorca, A. & Marín, O. Development and functional diversification of cortical interneurons. *Neuron* **100**, 294–313 (2018).

4. 4.

Harris, K. D. & Shepherd, G. M. G. The neocortical circuit: themes and variations. *Nat. Neurosci.* **18**, 170–181 (2015).

5. 5.

Oberst, P. et al. Temporal plasticity of apical progenitors in the developing mouse neocortex. *Nature* **573**, 370–374 (2019).

6. 6.

Mi, D. et al. Early emergence of cortical interneuron diversity in the mouse embryo. *Science* **360**, 81–85 (2018).

7. 7.

Nowakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

8. 8.

Mayer, C. et al. Clonally related forebrain interneurons disperse broadly across both functional areas and structural boundaries. *Neuron* **87**, 989–998 (2015).

9. 9.

Harwell, C. C. et al. Wide dispersion and diversity of clonally related inhibitory interneurons. *Neuron* **87**, 999–1007 (2015).

10. 10.

Fuentealba, L. C. et al. Embryonic origin of postnatal neural stem cells. *Cell* **161**, 1644–1655 (2015).

11. 11.
Golden, J. A., Fields-Berry, S. C. & Cepko, C. L. Construction and characterization of a highly complex retroviral library for lineage analysis. *Proc. Natl Acad. Sci. USA* **92**, 5704–5708 (1995).
12. 12.
Walsh, C. & Cepko, C. L. Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. *Science* **255**, 434–440 (1992).
13. 13.
Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* **21**, 410–427 (2020).
14. 14.
Kong, W. et al. CellTagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution. *Nat. Protoc.* **15**, 750–772 (2020).
15. 15.
Bowling, S. et al. An engineered CRISPR-Cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells. *Cell* **181**, 1410–1422.e27 (2020).
16. 16.
Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018).
17. 17.
Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
18. 18.
Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
19. 19.

Kalhor, R. et al. Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).

20. 20.

McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).

21. 21.

Cepko, C. L. et al. Lineage analysis with retroviral vectors. *Meth. Enzymol.* **327**, 118–145 (2000).

22. 22.

Biddy, B. A. et al. Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018).

23. 23.

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381 (2020).

24. 24.

Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).

25. 25.

Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).

26. 26.

Delgado, R. N. et al. Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature* <https://doi.org/10.1038/s41586-021-04230-7> (2021).

27. 27.

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502

(2015).

28. 28.

Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

29. 29.

Bayraktar, O. A. et al. Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell *in situ* transcriptomic map. *Nat. Neurosci.* **23**, 500–509 (2020).

30. 30.

Clavreul, S. et al. Cortical astrocytes develop in a plastic manner at both clonal and cellular levels. *Nat. Commun.* **10**, 4884 (2019).

31. 31.

Ge, W.-P., Miyawaki, A., Gage, F. H., Jan, Y. N. & Jan, L. Y. Local generation of glia is a major astrocyte source in postnatal cortex. *Nature* **484**, 376–380 (2012).

32. 32.

Qi, Y., Stapp, D. & Qiu, M. Origin and molecular specification of oligodendrocytes in the telencephalon. *Trends Neurosci.* **25**, 223–225 (2002).

33. 33.

Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

34. 34.

Ding, S. et al. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473–483 (2005).

35. 35.

Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

36. 36.

La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

37. 37.

Manning, C. S. et al. Quantitative single-cell live imaging links HES5 dynamics with cell-state and fate in murine neurogenesis. *Nat. Commun.* **10**, 2835 (2019).

38. 38.

Vitali, I. et al. Progenitor hyperpolarization regulates the sequential generation of neuronal subtypes in the developing neocortex. *Cell* **174**, 1264–1276.e15 (2018).

39. 39.

Wonders, C. P. et al. A spatial bias for the origins of interneuron subgroups within the medial ganglionic eminence. *Dev. Biol.* **314**, 127–136 (2008).

40. 40.

Flames, N. et al. Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J. Neurosci.* **27**, 9682–9695 (2007).

41. 41.

Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).

42. 42.

Rodin, R. E. et al. The landscape of mutational mosaicism in autistic and normal human cerebral cortex. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.02.11.944413> (2020).

43. 43.

Bae, T. et al. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* **359**, 550–555 (2018).

44. 44.

Ervony, G. D. et al. Integrated genome and transcriptome sequencing identifies a noncoding mutation in the genome replication factor DONSON as the cause of microcephaly-micromelia syndrome. *Genome Res.* **27**, 1323–1335 (2017).

45. 45.

Freed, D. & Pevsner, J. The contribution of mosaic variants to autism spectrum disorder. *PLoS Genet.* **12**, e1006245 (2016).

46. 46.

Lodato, M. A. et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).

47. 47.

Poduri, A., Evrony, G. D., Cai, X. & Walsh, C. A. Somatic mutation, genomic variation, and neurological disease. *Science* **341**, 1237758 (2013).

48. 48.

Daley, T. & Smith, A. D. Modeling genome coverage in single-cell sequencing. *Bioinformatics* **30**, 3159–3165 (2014).

49. 49.

Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

50. 50.

Zhao, L., Liu, Z., Levy, S. F. & Wu, S. Bartender: a fast and accurate clustering algorithm to count barcode reads. *Bioinformatics* **34**, 739–747 (2018).

51. 51.

Gaiano, N., Kohtz, J. D., Turnbull, D. H. & Fishell, G. A method for rapid gain-of-function studies in the mouse embryonic nervous system. *Nat. Neurosci.* **2**, 812–819 (1999).

52. 52.

Saito, T. In vivo electroporation in the embryonic mouse central nervous system. *Nat. Protoc.* **1**, 1552–1558 (2006).

53. 53.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

54. 54.

Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).

55. 55.

Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030.e16 (2018).

56. 56.

Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).

57. 57.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).

58. 58.

McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).

59. 59.

Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

60. 60.

Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).

Acknowledgements

We thank members of the Mayer laboratory, H. Baier, A. Borst, W. Denk, G. Evrony, A. Fabritius, M. Götz and R. Satija for feedback and discussion; J. Kuhl (somedonkey.com) for illustrations; M. Driessen and R. H. Kim from the MPIB Next Generation Sequencing core facility, I. Velasques and G. Eckstein from the Genomics Core facility at the Helmholtz Zentrum München (HMGU), M. Spitaler and M. Oster from the MPIB Imaging and FACS core facility, M. Gregory, C. Zollo and K. Ryan from the NYU FACS core facility, R. Kasper from the MPIN Imaging core facility, M. Fischer from the IT department and members of the MPIB/MPIN animal facility for their technical expertise. This work was supported by the Max-Planck Society, the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation program (ERC-2018-STG, grant agreement no. 803984) (to C.M.), the Simons Foundation (to G.F.), NIH 5R01NS081297-07 (to G.F.), 1UG3MH120096-01 (to G.F.), NIH 5R01MH071679-15 (to G.F.), RF1MH121268 from NIMH a Broad Center Regenerative Medicine & Stem Cell Research Innovation Award, a gift from the Bowes Foundation (to T.J.N.), generous anonymous donors, the Cure Alzheimer's Fund, The Blas Frangione Foundation, NIH R21NS111186 (to S.A.L.), NYU School of Medicine (to S.A.L. and R.C.B.), NIH F30MH114462 (to R.C.B.), T32GM007308 (to R.C.B.) and NIH P01NS074972 (to R.M.).

Funding

Open access funding provided by Max Planck Society.

Author information

Author notes

1. These authors contributed equally: Rachel C. Bandler, Ilaria Vitali, Ryan N. Delgado

Affiliations

1. Max Planck Institute of Neurobiology, Martinsried, Germany

Rachel C. Bandler, Ilaria Vitali, May C. Ho, Elena Dvoretskova, Josue S. Ibarra Molinas & Christian Mayer

2. NYU Neuroscience Institute, Langone Medical Center, New York, NY, USA

Rachel C. Bandler, Paul W. Frazel, Maesoumeh Mohammadkhani, Robert Machold & Shane A. Liddelow

3. Broad Institute, Stanley Center for Psychiatric Research, Cambridge, MA, USA

Rachel C. Bandler & Gord Fishell

4. Department of Anatomy, University of California, San Francisco, CA, USA

Ryan N. Delgado & Tomasz J. Nowakowski

5. Department of Psychiatry, University of California, San Francisco, CA, USA

Ryan N. Delgado & Tomasz J. Nowakowski

6. Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA, USA

Ryan N. Delgado & Tomasz J. Nowakowski

7. Max Planck Institute of Biochemistry, Martinsried, Germany

Sophia Maedler

8. Department of Neuroscience and Physiology, New York University Grossman School of Medicine, New York, NY, USA

Shane A. Liddelow

9. Department of Ophthalmology, New York University Grossman School of Medicine, New York, NY, USA

Shane A. Liddelow

10. Harvard Medical School, Department of Neurobiology, Boston, MA, USA

Gord Fishell

Contributions

R.C.B., G.F. and C.M. conceived the project. R.C.B., I.V., M.C.H., J.S.I.M., E.D. and C.M. designed the experiments and analysed the data. I.V., R.C.B. and C.M. led experimental work, assisted by E.D., S.M., P.W.F., M.M. and R.M. R.N.D. and T.J.N. developed STICR. M.C.H. and C.M. developed TrackerSeq. M.C.H., J.S.I.M., R.N.D. and C.M. implemented the computational analyses. R.C.B., I.V. and C.M prepared the manuscript, assisted by J.S.I.M., S.A.L. and G.F.

Corresponding author

Correspondence to [Christian Mayer](#).

Ethics declarations

Competing interests

S.A.L. is a founder of AstronauTx Ltd. All other authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Zoltan Molnar and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Extended data figures and tables

[Extended Data Fig. 1 Characterization of STICR datasets.](#)

a, Barcode diversity extrapolations derived from sequencing ca. 30 million reads of a representative STICR plasmid library. Mean \pm 95% CI. **b**, Images of whole mount brains injected with STICR. **c**, Representative FACS gating for negative control (matching non-injected brain) and STICR^{E10}. **d**, UMAP plot coloured and numbered for each unsupervised cluster. Each dot represents a single cell. **e**, UMAP plot coloured by cells with recovered transcriptome and cells with recovered transcriptome and lineage barcode. **f**, Examples of clones on UMAP plot. Sibling cell relationships represented by shapes (dot, square, star, cross, triangle). **g**, UMAP plot and bar graph showing the distribution of recovered cells, coloured by stage of injection. **h**, UMAP plot and bar graph showing the distribution of recovered cells, coloured by experimental replicate.

[Extended Data Fig. 2 Clonal relationships among cell classes.](#)

a, Single-cell heatmap showing the normalised expression of the top-5 marker genes for cell classes. Bars are colour-coded based on key in Fig. [1d](#). **b**, Immunohistochemistry showing STICR^{E10} viral labelling of excitatory

(CTIP2/CUX1) and inhibitory (GABA) neurons, astrocytes (S100B), oligodendrocytes (Olig2) and microglia (Iba1). White arrowheads indicate co-labelled cells, empty arrowheads indicate negative cells. Scale bar: 100 μ m (top), 30 μ m (insets), 50 μ m (bottom). Biological replicates, n = 2. **c**, Heatmap showing clonal distributions across cell classes for STICR^{E10} (left) and STICR^{E14} (right). Every horizontal line in the plot is a clone, with number of cells per clone indicated by colour. **d**, UpSet plots displaying clonal intersections between Astrocytes, Neurons, Oligodendrocyte Precursor Cells (OPCs), Neuronal precursors and Oligodentrocytes in STICR^{E10} (top) and STICR^{E14} (bottom). Bar graphs on top indicate the number of observed intersections. The bar graphs on the right depicts total cell number per cell class.

Extended Data Fig. 3 Differential gene expression analysis of cell type clusters.

Single-cell heatmap showing the normalised expression of the top-5 marker genes for cell types. Bars are colour-coded based on key in Fig. 2a. Astro, astrocyte; CGE, caudal ganglionic eminence; Ctx, neocortex; EAC, central extended amygdala; Excit, excitatory; Hip, hippocampus; IN, interneuron; Inhib, inhibitory; MGE, medial ganglionic eminence; NP, neuronal precursor; OB, olfactory bulb; Oligo, oligodendrocyte; OPC, oligodendrocyte precursor cell; PN, projection neuron; SPN, spiny projection neuron.

Extended Data Fig. 4 Spatial transcriptomic mapping of cell type clusters.

Spatial transcriptomic mapping showing localization of cell type clusters. Astro, astrocyte; CGE, caudal ganglionic eminence; Ctx, neocortex; EAC, central extended amygdala; Excit, excitatory; Hip, hippocampus; IN, interneuron; Inhib, inhibitory; MGE, medial ganglionic eminence; NP, neuronal precursor; OB, olfactory bulb; Oligo, oligodendrocyte; OPC, oligodendrocyte precursor cell; PN, projection neuron; SPN, spiny projection neuron.

Extended Data Fig. 5 Clonal coupling of forebrain cell types.

a, Heatmaps of STICR^{E10} (left) and STICR^{E14} (right) lineage coupling scores between pairs of cell types, clustered by correlation distance and linkage. Values range from positive (red, coupled) to negative (blue, anti-coupled). Light-grey lines link clusters across stages. Clonal groups are surrounded with grey lines and manually annotated (a-y). Astro, astrocyte; CGE, caudal ganglionic eminence; Ctx, neocortex; EAC, central extended amygdala; Excit, excitatory; Hip, hippocampus; IN, interneuron; Inhib, inhibitory; MGE, medial ganglionic eminence; NP, neuronal precursor; OB,

olfactory bulb; Oligo, oligodendrocyte; OPC, oligodendrocyte precursor cell; PN, projection neuron; SPN, spiny projection neuron. **b**, Feature plots highlighting examples of clonal groups that are maintained (left), split (middle) or merged (right) on the UMAP plot for STICR^{E10} and STICR^{E14}.

Extended Data Fig. 6 Excitatory and inhibitory lineages do not overlap in the murine cortex.

a, Violin plots showing the normalised marker gene expression Slc17a7 (marker gene for excitatory neurons) and Gad1 (marker gene for inhibitory neurons). CGE, caudal ganglionic eminence; Ctx, neocortex; Excit, excitatory; IN, interneuron; Inhib, inhibitory; MGE, medial ganglionic eminence. **b**, Spatial transcriptomic mapping showing localization of cortical neuron clusters in the mouse brain. **c**, Venn diagram quantifying the number of clones within excitatory and inhibitory clusters.

Extended Data Fig. 7 Clonal dispersion across cell types.

UpSet plots displaying intersections for cell type clusters in STICR^{E10} (top) and STICR^{E14} (bottom). Bar graphs on top indicate the number of observed intersections. The total cell number per cluster is shown on the right bar graphs. Intersection cut-off was set to 3 to fit on the plot. Astro, astrocyte; CGE, caudal ganglionic eminence; Ctx, neocortex; EAC, central extended amygdala; Excit, excitatory; Hip, hippocampus; IN, interneuron; Inhib, inhibitory; MGE, medial ganglionic eminence; NP, neuronal precursor; OB, olfactory bulb; Oligo, oligodendrocyte; OPC, oligodendrocyte precursor cell; PN, projection neuron; SPN, spiny projection neuron.

Extended Data Fig. 8 Lineage divergence in ventral inhibitory neurons.

a, Dot plot showing gene expression markers of ventral inhibitory neuron clusters. The size of the dots indicates the percentage of cells expressing a specific marker gene. The colour of the dots represents the average expression level. EAC, central extended amygdala; Inhib, inhibitory; MGE, medial ganglionic eminence; OB, olfactory bulb; PN, projection neuron; SPN, spiny projection neuron. **b**, Spatial transcriptomic mapping showing localization of ventral inhibitory neuron clusters in the mouse brain. **c**, Schematic representing the workflow for the manual dissection of striatum, olfactory bulb (OB), and amygdala from P6-P8 brains injected with STICR^{E14}. **d**, UpSet plot displaying clonal intersections amongst and within neuronal clusters of striatum, OB and amygdala. Only dispersing clones are shown. Bar graph on top indicate the number of observed intersections. The total cell number per cluster is represented on the right bar plot. Intersections amongst striatum, OB and amygdala are coloured in green. The heatmap shows the Spearman distance between the log-

normalized average cluster gene expression. The insets (top right) are UMAP plots of single cells from striatum, OB and amygdala datasets, with clusters coloured by cell class type. NP, neuronal precursor. **e**, Schematic of lineage divergence for ventral inhibitory neuron cell types.

Extended Data Fig. 9 Characterization of TrackerSeq.

a, Vector maps and cloning strategy of TrackerSeq. PBase: *piggyBac* transposase. **b**, Sanger sequencing results of individual *E. coli* colonies ('clones'), depicting the consensus sequence of the TrackerSeq lineage barcode. **c**, Pairwise hamming distance of 1000 barcodes randomly sampled from the TrackerSeq library. **d**, $\sim 3.6 \times 10^6$ raw sequencing reads were collapsed into $\sim 2 \times 10^5$ clusters, where each cluster is defined as a unique lineage barcode. **e**, UMAP plot of embryonic scRNA-seq datasets, cells coloured by dataset type (blue, TrackerSeq; grey, wild type). **f**, UMAP plot of single cells from the caudal, lateral and medial ganglionic eminences (CGE, LGE, MGE) of wild type (wt) embryos, coloured by clusters from Fig. [3c](#). **g**, Histogram showing distribution of clone sizes for TrackerSeq dataset. **h**, Clustered heatmap of TrackerSeq^{E12} barcodes. Rows are single GABAergic precursor cells for which both transcriptome and >1 TrackerSeq barcodes were retrieved; column represents unique TrackerSeq barcodes. Highlighted barcodes are those represented in Fig. [3i](#). **i**, Developmental trajectories of single-cell transcriptomes, coloured by pseudotime score. **j**, RNA Velocity plot. Arrows direction represent prediction of cells' future gene expression. **k**, Bar graph quantifying the correlation-based mapping of cells from the 5 embryonic precursor states to clusters of postnatal GABAergic forebrain neurons, including cortical interneuron types. EAC, central extended amygdala; IN, interneuron; Inhib, inhibitory; NP, neuronal precursor; OB, olfactory bulb; PN, projection neuron; SPN, spiny projection neuron. **l**, Scatter plots showing the normalized average cluster gene expression of the top 100 marker genes for a selected embryonic cluster and the top 100 marker genes for a selected postnatal cluster. Clusters were selected based on the mapping efficiency.

Extended Data Fig. 10 The contribution of lineage to the generation of cell diversity.

Schematic illustrating different scenarios of how cellular diversity could arise in the brain. Different cell types could arise from specified or fate-restricted progenitor cells (a), or be born sequentially from a common pool of progenitor cells (b, left). Such lineage-dependent processes suggest that cell-intrinsic mechanisms or local cues at the mitotic progenitor level, determine the fate of newborn cells. As another option, lineage-independent mechanisms such as extracellular induction, activity-dependent processes, or stochastic regulation (b, right) could drive the differentiation into different subtypes. Lineage-dependent and independent mechanisms are not mutually

exclusive. c) Convergence is the process by which similar cell states arise from different lineages.

Supplementary information

Reporting Summary

Supplementary Data 1

Descriptive report of collected datasets.

Supplementary Data 2

Top 20 differentially expressed genes for cell classes.

Supplementary Data 3

Top 20 differentially expressed genes for cell type clusters.

Supplementary Data 4

Top 20 differentially expressed genes for embryonic clusters.

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

Bandler, R.C., Vitali, I., Delgado, R.N. *et al.* Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature* **601**, 404–409 (2022).
<https://doi.org/10.1038/s41586-021-04237-0>

- Received: 15 January 2021
- Accepted: 12 November 2021
- Published: 15 December 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04237-0>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [**Individual human cortical progenitors can produce excitatory and inhibitory neurons**](#)
 - Ryan N. Delgado
 - Denise E. Allen
 - Tomasz J. Nowakowski

Nature (2021)

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04237-0>

- Article
- Open Access
- [Published: 18 November 2021](#)

Optimization of non-coding regions for a non-modified mRNA COVID-19 vaccine

- [Makda S. Gebre](#) [ORCID: orcid.org/0000-0001-8520-2178](#)¹,
[Susanne Rauch](#)²,
[Nicole Roth](#)²,
[Jingyou Yu](#)¹,
- [Abishek Chandrashekhar](#) [ORCID: orcid.org/0000-0001-7821-5552](#)¹,
- [Noe B. Mercado](#) [ORCID: orcid.org/0000-0001-7769-7326](#)¹,
- [Xuan He](#)¹,
- [Jinyan Liu](#)¹,
- [Katherine McMahan](#)¹,
- [Amanda Martinot](#) [ORCID: orcid.org/0000-0001-6237-6191](#)³,
- [David R. Martinez](#)⁴,
- [Victoria Giffin](#) [ORCID: orcid.org/0000-0002-9783-302X](#)¹,
- [David Hope](#) [ORCID: orcid.org/0000-0001-5654-1744](#)¹,
- [Shivani Patel](#) [ORCID: orcid.org/0000-0003-3475-1016](#)¹,
- [Daniel Sellers](#)¹,
- [Owen Sanborn](#)¹,
- [Julia Barrett](#)¹,
- [Xiaowen Liu](#) [ORCID: orcid.org/0000-0001-6030-8179](#)⁵,
- [Andrew C. Cole](#)⁵,
- [Laurent Pessant](#)⁶,
- [Daniel Valentin](#)⁶,
- [Zack Flinchbaugh](#)⁶,
- [Jake Valley-Ogunro](#)⁶,
- [Jeanne Muench](#)⁶,
- [Renita Brown](#)⁶,
- [Anthony Cook](#)⁶,
- [Elyse Teow](#)⁶,
- [Hanne Andersen](#) [ORCID: orcid.org/0000-0003-1103-9608](#)⁶,

- [Mark G. Lewis](#) [ORCID: orcid.org/0000-0001-7852-0135⁶](#),
- [Adrianus C. M. Boon](#) [ORCID: orcid.org/0000-0002-4700-8224⁷](#),
- [Ralph S. Baric](#) [ORCID: orcid.org/0000-0001-6827-8701⁴](#),
- [Stefan O. Mueller²](#),
- [Benjamin Petsch^{2 na1}](#) &
- [Dan H. Barouch](#) [ORCID: orcid.org/0000-0001-5127-4659^{1,8 na1}](#)

[*Nature*](#) volume **601**, pages 410–414 (2022)

- 22k Accesses
- 2 Citations
- 126 Altmetric
- [Metrics details](#)

Subjects

- [RNA vaccines](#)
- [SARS-CoV-2](#)

Abstract

The CVnCoV (CureVac) mRNA vaccine for severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) was recently evaluated in a phase 2b/3 efficacy trial in humans¹. CV2CoV is a second-generation mRNA vaccine containing non-modified nucleosides but with optimized non-coding regions and enhanced antigen expression. Here we report the results of a head-to-head comparison of the immunogenicity and protective efficacy of CVnCoV and CV2CoV in non-human primates. We immunized 18 cynomolgus macaques with two doses of 12 µg lipid nanoparticle-formulated CVnCoV or CV2CoV or with sham ($n = 6$ per group). Compared with CVnCoV, CV2CoV induced substantially higher titres of binding and neutralizing antibodies, memory B cell responses and T cell responses as well as more potent neutralizing antibody responses against SARS-CoV-2 variants, including the Delta variant. Moreover, CV2CoV was found to be comparably immunogenic to the BNT162b2 (Pfizer) vaccine in macaques. Although CVnCoV provided partial protection against SARS-CoV-2 challenge, CV2CoV afforded more robust protection with markedly lower viral loads in the upper and lower respiratory tracts. Binding and neutralizing antibody titres were correlated with protective efficacy. These data demonstrate that optimization of non-coding regions can greatly improve the immunogenicity and

protective efficacy of a non-modified mRNA SARS-CoV-2 vaccine in non-human primates.

[Download PDF](#)

Main

Efficacy results in humans have recently been reported for the CVnCoV (CureVac) mRNA vaccine in the phase 2b/3 HERALD trial in a population that included multiple viral variants. In this trial, the observed vaccine efficacy against symptomatic coronavirus disease 2019 (COVID-19) was approximately 48% and 53% in the overall study population and in a subgroup of participants 18–60 years of age, respectively¹. CV2CoV is a second-generation mRNA vaccine that incorporates modifications of non-coding regions that were selected by empiric screening for improved antigen expression^{2,3}. Both CVnCoV and CV2CoV are based on RNAactive technology^{4,5,6,7} and consist of non-chemically modified sequence-engineered mRNA without pseudouridine^{6,7,8,9,10,11,12}. Both vaccines encode the same full-length, pre-fusion stabilized severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) spike protein^{13,14} and are encapsulated in lipid nanoparticles (LNPs) with identical composition. CV2CoV has been engineered with different non-coding regions flanking the open reading frame, which have previously been shown to improve transgene expression³ and protection against SARS-CoV-2 in ACE2-transgenic mice². Specifically, CV2CoV includes 5' untranslated region (UTR) HSD17B4 and 3' UTR PSMB3 elements followed by a histone stem-loop motif and a poly(A) sequence (Fig. 1 and [Methods](#)). In the present study, we make a head-to-head comparison of the immunogenicity and protective efficacy of CVnCoV and CV2CoV against SARS-CoV-2 challenge in non-human primates.

Fig. 1: Vaccine design and study schema.

 **figure 1**

a, Designs of the CVnCoV and CV2CoV mRNA vaccine candidates. Both vaccines are based on CureVac's RNAActive platform and encode SARS-CoV-2 spike protein with di-proline mutations. The vaccines differ in their unique non-coding regions, as shown. **b**, Non-human primate vaccine study schema. Cynomolgus macaques were immunized intramuscularly (i.m.) on day (D) 0 with CVnCoV ($n = 6$) or CV2CoV ($n = 6$) mRNA vaccine or were designated as sham ($n = 6$). The animals were boosted at week 4 and were challenged at week (W) 8. Samples were collected weekly after immunization and on days 0, 1, 2, 4, 7 and 10 after challenge for immunological and virological assays. PP, K986P and V987P mutations; HSL, histone stem-loop.

Vaccine immunogenicity

We immunized 18 cynomolgus macaques intramuscularly with 12 µg CVnCoV, 12 µg CV2CoV or sham vaccine (Fig. 1b). The animals were primed at week 0 and were boosted at week 4. No clinical adverse effects were observed following vaccination. To assess innate immune responses, sera were isolated from all animals 24 h after the first vaccination to evaluate innate cytokine responses. CV2CoV induced higher levels of IFN α 2a, IP-10 and MIP-1 than CVnCoV ($P = 0.0152$, $P = 0.0152$ and $P = 0.0411$, respectively; Extended Data Fig. 1).

Binding antibody responses were assessed by performing receptor-binding domain (RBD)-specific enzyme-linked immunosorbent assays (ELISAs) at multiple time points following immunization^{15,16}. At week 2, binding antibody titres were detected only with CV2CoV and not with CVnCoV, with median values of 25 (range, 25–25)

and 799 (range, 82–2,010) for CVnCoV and CV2CoV, respectively (Fig. 2a). One week after the week-4 boost, the antibody titres were increased in both groups, with medians of 48 (range, 75–710) and 28,407 (range, 2,714–86,541) for CVnCoV and CV2CoV, respectively (Fig. 2a). By week 8, the binding antibody titres had increased in the CVnCoV group but were still >50 times lower than those in the CV2CoV group ($P = 0.0043$), with median values of 214 (range, 47–1,238) and 14,827 (range, 2,133–37,079), respectively.

Fig. 2: CV2CoV elicits high levels of binding and neutralizing antibody responses in macaques.

 figure 2



Animals ($n = 6$ per group) were vaccinated twice with 12 μg of CVnCoV or CV2CoV on day 0 and on day 28 or remained untreated as negative controls (sham). **a, b**, Titres of RBD-binding antibodies (**a**) and pseudovirus neutralizing antibodies (NAb) against the ancestral SARS-CoV-2 strain (**b**) were evaluated at different time points after the first (weeks 0, 1, 2 and 4) and second (weeks 5, 6 and 8) vaccinations. **c, d**, Sera collected on day 42 (week 6) were analysed for pseudovirus (**c**) and live-virus (**d**) neutralizing antibody titres against virus with the D614G mutation and the B.1.1.7 (Alpha), B.1.351 (Beta) and B.1.617.2 (Delta) variants. **e**, Sera collected from non-human primates immunized with 12 μg of CVnCoV or 30 μg of BNT162b2 on day 35 (week 5) after boosting were analysed for pseudovirus neutralizing antibody titres against the ancestral WA/2020 (WT) strain. Each dot represents an individual animal and bars depict the median; the dashed line shows the limit of detection.

Source data

Neutralizing antibody responses were assessed by pseudovirus neutralization assay using the vaccine-matched SARS-CoV-2 wild-type (WT) WA1/2020 strain^{15,16,17}. The neutralizing antibody titres followed a trend similar to that of the binding antibody titres (Fig. 2b). At week 2, neutralizing antibodies were detected only with CV2CoV

and not with CVnCoV, with median values of 20 (range, 20–20) and 131 (range, 62–578) for CVnCoV and CV2CoV, respectively (Fig. 2b). One week after the week 4 boost, the neutralizing antibody titres were increased, with median values of 55 (range, 20–302) and 15,827 (range, 3,985–81,081) for CVnCoV and CV2CoV, respectively. By week 8, the neutralizing antibody titres had increased in the CVnCoV group but were still >20 times lower than those in the CV2CoV group ($P = 0.0022$), with median values of 196 (range, 20–405) and 4,752 (range, 414–6,793), respectively.

At week 6, the median pseudovirus neutralizing antibody titres against the D614G, B.1.1.7 (Alpha) and B.1.351 (Beta) variants for CVnCoV were 121, 101 and 189, respectively, while they were 4,962, 1,813 and 755 for CV2CoV (Fig. 2c). The median pseudovirus neutralizing antibody titres against the C.37 (Lambda), B.1.617.1 (Kappa) and B.1.617.2 (Delta) variants for CVnCoV were 516, 158 and 36, respectively, while they were 1,195, 541 and 568 for CV2CoV (Extended Data Fig. 2). The pseudovirus neutralizing antibody titres induced by CV2CoV were higher than those induced by CVnCoV for the WT (WA1/2020), D614G, B.1.1.7 (Alpha), B.1.351 (Beta), C.37 (Lambda), B.1.617.1 (Kappa) and B.1.617.2 (Delta) strains ($P = 0.0043$, 0.0087, 0.0043, 0.1320, 0.026, 0.0022 and 0.0043, respectively). Taken together, these data show that CV2CoV induces substantially higher pseudovirus neutralizing antibody titres against SARS-CoV-2 variants than CVnCoV.

Live-virus neutralizing antibody titres¹⁸ were largely consistent with those for the pseudovirus. The live-virus neutralizing antibody responses elicited by CV2CoV were higher than those elicited by CVnCoV against the WA1/2020 and B.1.617.2 (Delta) strains ($P = 0.0466$ and 0.0152, respectively), with similar trends for B.1.1.7 (Alpha) and B.1.351 (Beta) ($P = 0.0628$ and 0.1450, respectively) (Fig. 2d).

We also compared the pseudovirus neutralizing antibody titres induced in macaques by two immunizations with 12 µg of CV2CoV to those induced by two immunizations with 30 µg of the Pfizer BNT162b2 clinical vaccine obtained as leftover product from pharmacies. At peak immunity at week 5, the neutralizing antibody responses induced by CV2CoV were comparable to those induced by BNT162b2 (Fig. 2e).

Most SARS-CoV-2 RBD-specific B cells reside within the memory B cell pool¹⁹. We used flow cytometry to assess memory B cell responses in the blood of non-human primates vaccinated with CVnCoV, CV2CoV or sham²⁰. Higher numbers of RBD-specific and spike-specific memory B cells were detected in the CV2CoV-vaccinated animals as compared with those vaccinated with CVnCoV at week 6 ($P = 0.022$ and $P = 0.0152$, respectively) (Extended Data Fig. 3a,b). T cell responses were assessed by interferon γ (IFN γ) and interleukin (IL)-4 enzyme-linked immunosorbent spot (ELISPOT) assay using pooled spike peptides at week 6. IFN γ responses were detected in both groups but were higher in the CV2CoV group ($P = 0.0065$) (Extended

Data Fig. 3c). IL-4 responses were not detectable, suggesting that CVnCoV and CV2CoV induce T helper type 1-biased responses (Extended Data Fig. 3d).

Protective efficacy

All animals were challenged at week 8 with 1.0×10^5 median tissue culture infectious doses (TCID₅₀) of the SARS-CoV-2 WA1/2020 strain via the intranasal and intratracheal routes. Viral loads were assessed in bronchoalveolar lavage (BAL) and nasal swab samples collected on days 1, 2, 4, 7 and 10 following challenge by quantitative PCR with reverse transcription (RT-PCR) specific for subgenomic RNA (sgRNA)²¹. The sgRNA levels in the BAL and nasal swab samples in the sham group peaked on day 2 and largely resolved by day 10. The sham controls had peak medians of 6.02 (range, 4.62–6.81) log₁₀-transformed sgRNA copies per ml in the BAL and 7.35 (range, 5.84–8.09) log₁₀-transformed sgRNA copies per swab in the nasal swab samples on day 2 (Fig. 3). The CVnCoV-immunized animals showed peak medians of 4.92 (range, 2.40–6.61) log₁₀-transformed sgRNA copies per ml in the BAL and 6.42 (range, 4.46–7.81) log₁₀-transformed sgRNA copies per swab in the nasal swab samples (Fig. 3). The CV2CoV-immunized animals exhibited peak medians of 2.90 (range, 1.70–4.64) log₁₀-transformed sgRNA copies per ml in the BAL and 3.17 (range, 2.59–5.63) log₁₀-transformed sgRNA copies per swab in the nasal swab samples (Fig. 3), with resolution of sgRNA levels in the BAL samples by day 2 in most animals and by day 4 in all animals. Overall, CV2CoV resulted in significantly lower peak viral loads than CVnCoV in both the BAL ($P = 0.0411$) and nasal swab ($P = 0.0087$) samples (Fig. 4a,b).

Fig. 3: Protective efficacy of CV2CoV.

 **figure 3**

Negative-control animals (sham) and animals ($n = 6$ per group) vaccinated on day 0 and day 28 with 12 µg of CVnCoV or CV2CoV were challenged with 1.0×10^5 TCID₅₀ of SARS-CoV-2 (USA-WA1/2020) via the intranasal and intratracheal routes.

a, b, BAL (**a**) and nasal swab (**b**) samples collected on days 1, 2, 4, 7 and 10 after challenge were analysed for levels of replicating virus by RT-PCR specific for sgRNA. Thin black lines represent individual animals and thick red lines depict the median; the dashed line shows the limit of detection.

[Source data](#)

Fig. 4: Titres of binding and neutralizing antibodies elicited following CVnCoV and CV2CoV vaccination ($n = 6$ per group) correlate with protection against SARS-CoV-2.

 **figure 4**

a, b, Summary of peak viral loads following SARS-CoV-2 challenge in BAL and nasal swab (NS) samples. Animals were challenged with 1.0×10^5 TCID₅₀ of SARS-CoV-2 derived from strain USA-WA1/2020 (NR-52281, BEI Resources). **c–f**, Antibody correlates of protection for binding antibodies (**c, d**) and neutralizing antibodies (**e, f**). Statistical analysis was performed using the two-tailed non-parametric Mann–Whitney test, and correlation was analysed by two-sided Spearman rank-correlation test. The bars indicate median values.

Source data

We next evaluated the immune correlates of protection. The log₁₀-transformed ELISA and neutralizing antibody titres at week 6 were inversely correlated with the peak log₁₀-transformed sgRNA copies per ml in the BAL samples ($P = 0.0008$, $R = -0.7148$ and $P = 0.0015$, $R = -0.6912$, respectively, by two-sided Spearman rank-correlation test) (Fig. 4c, e) and with the peak sgRNA copies per nasal swab in the nasal swab samples ($P < 0.0001$, $R = -0.8346$ and $P < 0.0001$, $R = -0.8766$, respectively, by two-sided Spearman rank-correlation test) (Fig. 4d, f). Consistent with prior observations from our laboratory and others^{15,16,22}, these findings suggest that binding and neutralizing antibody titres are important correlates of protection for these SARS-CoV-2 vaccines in non-human primates. Similar correlates of protection were observed with viral loads assessed as area under the curve (Extended Data Fig. 4). Moreover, we assessed infectious virus titres by TCID₅₀ assay on day 2 after challenge, which showed no detectable virus in five of six animals in the CV2CoV group (Extended Data Fig. 5).

Following challenge, we observed anamnestic binding and neutralizing antibody responses in all CVnCoV-vaccinated animals and in a subset of the CV2CoV-vaccinated animals¹⁶ (Extended Data Fig. 6). On day 10 after challenge, the animals were necropsied, and their lung tissues were evaluated by histopathology. Viral replication was largely resolved by day 10 in the animals vaccinated with CVnCoV and CV2CoV, and those with sham treatment had higher cumulative lung pathology scores¹⁹ (CVnCoV animals compared with sham controls, $P = 0.0368$; CV2CoV animals compared with sham controls, $P = 0.0022$) (Extended Data Fig. 7a). Animals in the sham group also had more lung lobes affected (Extended Data Fig. 7b) and more extensive lung lesions, with a greater proportion of lung lobes showing evidence of interstitial inflammation, alveolar inflammatory infiltrates and type II pneumocyte hyperplasia (Extended Data Fig. 7c–h). No significant eosinophilia was observed. The pathological lesions in vaccinated animals were similar to those observed for animals in the sham group (Extended Data Fig. 7i–l) but were overall fewer in number and more focal in distribution.

Discussion

CV2CoV elicited substantially greater humoral and cellular immune responses and provided significantly improved protective efficacy against SARS-CoV-2 challenge as compared with CVnCoV in macaques. These data suggest that optimization of non-coding elements of the mRNA backbone can substantially improve the immunogenicity and protective efficacy of mRNA vaccines. Both CVnCoV and CV2CoV contain only non-modified nucleosides with no pseudouridine or derivates, and CV2CoV has previously been shown to lead to higher antigen expression than CVnCoV in cell culture³. The neutralizing antibody titres induced by CV2CoV were comparable in macaques to those induced by the clinical BNT162b2 vaccine, which incorporates pseudouridine. These results suggest that strategies other than nucleoside modification can also markedly improve mRNA potency.

Previous studies with rodents and non-human primates have demonstrated protection by CVnCoV^{2,23,24}. However, protection in macaques was primarily observed in the lower respiratory tract^{23,24}. In the present study, CVnCoV provided only modest viral load reductions in BAL and nasal swab samples compared with sham controls. In contrast to CVnCoV, CV2CoV induced >10-fold-higher neutralizing antibody responses against multiple viral variants and provided >3 log reductions in sgRNA copies per ml in BAL and >4 log reductions in sgRNA copies per swab in nasal swab samples compared with sham controls.

Previous mRNA vaccine clinical trials have demonstrated onset of protective efficacy after the first dose with improved protection after the boost immunization^{25,26}. In the

present study, the prime immunization with CV2CoV induced binding and neutralizing antibodies in all macaques by week 2, and these responses had increased substantially by 1 week after the boost immunization. The neutralizing antibody titres induced by CV2CoV in this study also appear to be similar to those reported for other mRNA vaccines in macaques^{27,28}. Moreover, the neutralizing antibody titres induced by BNT162b2 in our study (Fig. 2e) were comparable to those reported for BNT162b2 in a prior study²⁸.

As previously reported for other vaccines^{29,30,31,32,33}, the neutralizing antibody titres against certain SARS-CoV-2 variants, such as the B.1.351 (Beta) and B.1.617.2 (Delta) variants, were lower than those against the parental strain WA1/2020. Although our challenge virus in this study was SARS-CoV-2 WA1/2020, the neutralizing antibody titres elicited by CV2CoV to viral variants exceeded the values we previously reported as threshold titres for protection (50–100)^{17,19,22}. However, future studies will be required to directly assess the protective efficacy of CV2CoV against SARS-CoV-2 variants of concern in non-human primates.

CV2CoV induced both antigen-specific memory B cell responses and T cell responses. Although the correlates of protection in this study were binding and neutralizing antibody titres^{34,35}, it is likely that CD8⁺ T cells contribute to viral clearance in tissues^{36,37}. We previously reported that depletion of CD8⁺ T cells partially abrogated protective efficacy against SARS-CoV-2 re-challenge in convalescent macaques²². Memory B cells might contribute to the durability of antibody responses^{38,39}; B cell germinal centre responses and the durability of protective efficacy following CV2CoV vaccination remain to be determined. Moreover, although this study was not specifically designed as a safety study, it is worth noting that we did not observe any adverse effects following CVnCoV or CV2CoV vaccination, nor did we observe unexpected or enhanced pathology in the vaccinated animals at necropsy⁴⁰.

In summary, our data show that optimization of non-coding regions in a SARS-CoV-2 mRNA vaccine can substantially improve its immunogenicity against multiple viral variants and can enhance its protective efficacy against SARS-CoV-2 challenge in macaques. The improved characteristics of CV2CoV over those of CVnCoV might translate into increased efficacy in humans; accordingly, clinical trials of CV2CoV are planned.

Methods

mRNA vaccines

The two mRNA vaccines, CVnCoV and CV2CoV, are based on CureVac's RNActive platform (claimed and described in, for example, patents WO2002098443 and WO2012019780) and include no chemically modified nucleosides. They are composed of a 5' cap1 structure, a G+C-enriched open reading frame, a 3' UTR and a vector-encoded poly(A) stretch. CVnCoV contains a cleanCap (Trilink) and parts of the 3' UTR of the *Homo sapiens* alpha-haemoglobin gene as the 3' UTR, followed by a poly(A)₆₄ stretch, a poly(C)₃₀ stretch and a histone stem-loop^{22,23}. CV2CoV has previously been described and contains a cleanCap followed by the 5' UTR from the human hydroxysteroid 17-beta dehydrogenase 4 gene (*HSD17B4*) and a 3' UTR from the human proteasome 20S subunit beta 3 gene (*PSMB3*) followed by a histone stem-loop and a poly(A)₁₀₀ stretch³. The constructs were encapsulated in LNP by Acuitas Therapeutics (CV2CoV) or Polymun Scientific Immunbiologische Forschung (CVnCoV). LNPs are composed of ionizable amino lipid, phospholipid and cholesterol and PEGylated lipid; the compositions for CVnCoV and CV2CoV are identical. Both mRNAs encode SARS-CoV-2 full-length spike protein containing stabilizing K986P and V987P mutations (NCBI reference sequence NC045512.2).

Animals and study design

Eighteen cynomolgus macaques of both sexes between the ages of 3 and 20 years were randomly assigned to three groups. The animals received either CVnCoV ($n = 6$) or CV2CoV ($n = 6$) mRNA vaccine or were designated as sham controls ($n = 6$). The mRNA vaccines were administered intramuscularly at a 12- μg dose in the left quadriceps on day 0. Boost immunizations were similarly administered at week 4. At week 8, all animals were challenged with 1.0×10^5 TCID₅₀ of SARS-CoV-2 derived from USA-WA1/2020 (NR-52281, BEI Resources)¹⁷. The challenge virus was administered as 1 ml by the intranasal route (0.5 ml in each naris) and 1 ml by the intratracheal route. All animals were killed 10 d after challenge. Immunological and virological assays were performed with blinding. All animals were housed at Bioqual. All animal studies were conducted in compliance with all relevant local, state and federal regulations and were approved by the Bioqual Institutional Animal Care and Use Committee.

Cytokine analyses

The serum levels of 19 analytes that have been associated with immune response to viral infection were tested using the U-PLEX Viral Combo 1 (NHP) kit (K15069L-1) obtained from Meso Scale Discovery. The 19 analytes and their detection limits (LLODs) included G-CSF (1.5 pg ml⁻¹), GM-CSF (0.12 pg ml⁻¹), IFN α 2a (1.7 pg ml⁻¹), IFN γ (1.7 pg ml⁻¹), IL-1RA (1.7 pg ml⁻¹), IL-1 β (0.15 pg ml⁻¹), IL-4 (0.06 pg ml⁻¹), IL-5 (0.24 pg ml⁻¹), IL-6 (0.33 pg ml⁻¹), IL-7 (1.5 pg ml⁻¹), IL-8 (0.15 pg ml⁻¹),

IL-9 (0.14 pg ml⁻¹), IL-10 (0.14 pg ml⁻¹), IL-12p70 (0.54 pg ml⁻¹), IP-10 (0.49 pg ml⁻¹), MCP-1 (0.74 pg ml⁻¹), MIP-1 α (7.7 pg ml⁻¹), TNF (0.54 pg ml⁻¹) and VEGF-A (2.0 pg ml⁻¹). All serum samples were assayed in duplicate. The assay was performed by the Metabolism and Mitochondrial Research Core (Beth Israel Deaconess Medical Center, Boston, MA) following the manufacturer's instructions. The assay plates were read by a MESO QuickPlex SQ 120 instrument, and the data were analysed using Discovery Workbench 4.0 software.

ELISA

RBD-specific binding antibodies were assessed by ELISA as described previously^{16,17}. In brief, 96-well plates were coated with 1 μ g ml⁻¹ SARS-CoV-2 RBD protein (40592-VNAH, Sino Biological) in 1 \times DPBS and were incubated at 4 °C overnight. After incubation, the plates were washed once with wash buffer (0.05% Tween-20 in 1 \times DPBS) and were blocked with 350 μ l casein block per well for 2–3 h at room temperature. After incubation, the block solution was discarded, and the plates were blotted dry. Serial dilutions of heat-inactivated serum diluted in casein block were added to the wells, and the plates were incubated for 1 h at room temperature. Next, the plates were washed three times and were then incubated for 1 h with a 1:1,000 dilution of anti-macaque IgG HRP (NIH NHP Reagent Program) at room temperature in the dark. The plates were then washed three more times, and 100 μ l of SeraCare KPL TMB SureBlue Start solution was added to each well; plate development was halted by the addition of 100 μ l of SeraCare KPL TMB Stop solution per well. The absorbance at 450 nm was recorded using a VersaMax or Omega microplate reader. The ELISA endpoint titres were defined as the highest reciprocal serum dilution that yielded an absorbance >0.2, and the log₁₀ endpoint titres are reported. The immunological assays were performed with blinding.

Pseudovirus neutralization assay

SARS-CoV-2 pseudoviruses encoding a luciferase reporter gene were generated as described previously¹⁵. In brief, the packaging plasmid psPAX2 (AIDS Resource and Reagent Program), luciferase reporter plasmid pLenti-CMV Puro-Luc (Addgene) and spike protein-encoding pcDNA3.1-SARS CoV-2 S Δ CT plasmid of variants were co-transfected into HEK293T cells by Lipofectamine 2000 (ThermoFisher Scientific). Pseudoviruses of SARS-CoV-2 variants were generated by using the WA1/2020 strain (Wuhan/WIV04/2019; GISAID accession ID, EPI_ISL_402124), the strain with a D614G mutation, the B.1.1.7 variant (GISAID accession ID, EPI_ISL_601443), the B.1.351 variant (GISAID accession ID, EPI_ISL_712096), the C37 variant (GenBank ID, QRX62290), the B.1.671.1 variant (GISAID accession ID, EPI_ISL_1384866) and the B.1.617.2 variant (GISAID accession ID, EPI_ISL_2020950). Supernatants

containing the pseudotype viruses, which were purified by centrifugation and filtration with a 0.45- μ m filter, were collected 48 h after transfection. To determine the neutralization activity of the plasma or serum samples from the animals studied, HEK293T-hACE2 cells were seeded in 96-well tissue culture plates at a density of 1.75×10^4 cells per well overnight. Threefold serial dilutions of heat-inactivated serum or plasma samples were prepared and mixed with 50 μ l of pseudovirus. The mixture was incubated at 37 °C for 1 h before being added to the HEK293T-hACE2 cells. The cells were lysed 48 h after infection in Steady-Glo Luciferase Assay buffer (Promega) according to the manufacturer's instructions. The SARS-CoV-2 neutralization titres were defined as the sample dilution at which a 50% reduction in relative light units (RLU) was observed relative to the average of the virus control wells.

Live-virus neutralization assay

Full-length SARS-CoV-2 WA1/2020, B.1.1.7, B.1.351 and B.1.617.2 viruses were designed to encode nanoluciferase (nLuc) and were recovered via reverse genetics¹⁸. One day before the assay, Vero E6 USAMRID cells were plated at 20,000 cells per well in clear-bottomed, black-walled plates. The cells were inspected to ensure confluence on the day of the assay. The serum samples were tested at a starting dilution of 1:20 and were serially diluted threefold for up to nine dilution spots. The serially diluted serum samples were mixed with diluted virus in an equal volume. The antibody–virus and virus-only mixtures were then incubated at 37 °C with 5% CO₂ for 1 h. After incubation, the serially diluted sera and virus-only controls were added in duplicate to the cells at 75 plaque-forming units at 37 °C with 5% CO₂. The cells were lysed 24 h later, and the luciferase activity was measured via Nano-Glo Luciferase Assay System (Promega) according to the manufacturer's specifications. The luminescence was measured by a Spectramax M3 plate reader (Molecular Devices). Virus neutralization titres were defined as the sample dilution at which a 50% reduction in RLU was observed relative to the average of the virus control wells.

B cell immunophenotyping

Fresh peripheral blood mononuclear cells were stained with Aqua live/dead dye (Invitrogen) for 20 min, washed with 2% FBS in DPBS and suspended in 2% FBS in DPBS with Fc Block (BD) for 10 min, followed by staining with monoclonal antibodies against CD45 (clone D058-1283, BUV805), CD3 (clone SP34.2, APC-Cy7), CD7 (clone M-T701, Alexa700), CD123 (clone 6H6, Alexa700), CD11c (clone 3.9, Alexa700), CD20 (clone 2H7, PE-Cy5), IgA (goat polyclonal antibodies, APC), IgG (clone G18-145, BUV737), IgM (clone G20-127, BUV396), IgD (goat polyclonal antibodies, PE), CD80 (clone L307.4, BV786), CD95 (clone DX2, BV711), CD27 (clone M-T271, BUV563), CD21 (clone B-ly4, BV605), CD14 (clone M5E2, BV570)

and CD138 (clone DL-101, PE-CF594). The cells were also stained for SARS-CoV-2 antigens including biotinylated SARS-CoV-2 RBD protein (Sino Biological) and full-length SARS-CoV-2 spike protein (Sino Biological) labelled with FITC and DyLight 405 (DyLight 405 Conjugation Kit and FITC Conjugation Kit, Abcam) at 4 °C for 30 min. After staining, the cells were washed twice with 2% FBS in DPBS, incubated with BV650 streptavidin (BD Pharmingen) for 10 min and then washed twice with 2% FBS in DPBS. After staining, the cells were washed and fixed with 2% paraformaldehyde. All data were acquired on a BD FACSymphony flow cytometer. Subsequent analyses were performed using FlowJo software (Treestar, v.9.9.6). The immunological assays were performed with blinding.

IFN γ ELISPOT assay

ELISPOT plates were coated with mouse anti-human IFN γ monoclonal antibody (BD Pharmingen) at a concentration of 5 μ g per well overnight at 4 °C. The plates were washed with DPBS containing 0.25% Tween-20 and were blocked with R10 medium (RPMI with 11% FBS and 1.1% penicillin–streptomycin) for 1 h at 37 °C. The S1 and S2 peptide pools (custom made, JPT Peptide Technologies) used in the assay contained peptides of 15 amino acids in length, overlapping by 11 amino acids, that spanned the protein sequence and reflect the N-terminal and C-terminal halves of the protein, respectively. The S1 and S2 peptide pools were prepared at a concentration of 2 μ g per well, and 200,000 cells per well were added. The peptides and cells were incubated for 18–24 h at 37 °C. All steps following this incubation were performed at room temperature. The plates were washed with ELISPOT wash buffer and were incubated for 2 h with 1 μ g ml $^{-1}$ rabbit polyclonal anti-human IFN γ biotin obtained from U-Cytech. The plates were washed a second time and were then incubated for 2 h with 1 μ g ml $^{-1}$ streptavidin–alkaline phosphatase antibody obtained from Southern Biotech. The final wash was followed by the addition of nitro-blue tetrazolium chloride/5-bromo-4-chloro 3' indolyl phosphate *p*-toluidine salt (NBT/BCIP chromagen) substrate solution (ThermoFisher Scientific) for 7 min. The chromogen was discarded, and the plates were washed with water and were dried in a dim location for 24 h. The plates were then scanned and counted using an ELISPOT analyser (Immunospot).

IL-4 ELISPOT assay

ELISPOT plates precoated with monoclonal antibody against IL-4 (Mabtech) were washed and blocked. The assay was then performed as described above except that the development time with NBT/BCIP chromagen substrate solution was 12 min.

Subgenomic RT–PCR assay

SARS-CoV-2 *E* gene sgRNA was assessed by RT–PCR using primers and probes as previously described^{15,17}. A standard was generated by first synthesizing a gene fragment of the subgenomic *E* gene. The gene fragment was subsequently cloned into the pcDNA3.1+ expression plasmid using restriction site cloning (Integrated DNA Technologies). The insert was transcribed in vitro to RNA using the AmpliCap-Max T7 High Yield Message Maker kit (CellScript). Log dilutions of the standard were prepared for RT–PCR assays, ranging from 1×10^{10} copies to 1×10^{-1} copies. The viral loads were quantified from BAL fluid and nasal swab samples. RNA extraction was performed on a QIAcube HT using the IndiSpin QIAcube HT Pathogen kit according to the manufacturer’s specifications (Qiagen). The standard dilutions and extracted RNA samples were reverse-transcribed using SuperScript VILO Master Mix (Invitrogen) following the cycling conditions described by the manufacturer. A Taqman custom gene expression assay (ThermoFisher Scientific) was designed using the sequences targeting the *E* gene sgRNA. The sequences for the custom assay were as follows: forward primer, sgLeadCoV2.Fwd: 5'-CGATCTCTGTAGATCTGTTCTC-3'; E_Sarbeco_R: 5'-ATATTGCAGCAGTACCGCACACA-3'; E_Sarbeco_P1 (probe): 5'-VIC-ACACTAGCCATCCTTACTGCGCTTCG-MGBNFQ-3'. Reactions were carried out in duplicate for samples and standards on QuantStudio 6 and 7 Flex Real-Time PCR systems (Applied Biosystems) with the following thermal cycling conditions: initial denaturation at 95 °C for 20 s followed by 45 cycles of 95 °C for 1 s and 60 °C for 20 s. Standard curves were used to calculate the sgRNA copies per millilitre or per swab. The quantitative assay sensitivity was determined as 50 copies per millilitre or per swab.

TCID₅₀ assay

Vero TMPRSS2 cells (obtained from A. Creanga, NIH) were plated at 25,000 cells per well in DMEM with 10% FBS and gentamicin, and the cultures were incubated at 37 °C, 5.0% CO₂. Medium was aspirated and replaced with 180 µl of DMEM with 2% FBS and gentamicin. Serial dilution of samples as well as positive (virus stock of known infectious titre) and negative (medium only) controls were included in each assay. The plates were incubated at 37 °C, 5.0% CO₂, for 4 d, and the cell monolayers were visually inspected for cytopathic effects. TCID₅₀ was calculated using the Read–Muench formula.

Histopathology

At the time of fixation, lungs were suffused with 10% formalin to expand the alveoli. All tissues were fixed in 10% formalin and block-sectioned at 5 µm. The slides were baked for 30–60 min at 65 °C, deparaffinized in xylene, rehydrated through a series of

graded ethanol to distilled water and then stained with haematoxylin and eosin. Blinded histopathological evaluation was performed by a board-certified veterinary pathologist (A.J.M.).

Statistical analyses

Statistical analyses were performed using GraphPad Prism (version 9.0) software (GraphPad Software), and comparisons between groups were performed using a two-tailed non-parametric Mann–Whitney *U* test. *P* values of less than 0.05 were considered as significant. Correlations were assessed by applying two-sided Spearman rank-correlation tests.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

All data are available in the manuscript and its [Supplementary Information](#). [Source data](#) are provided with this paper.

References

1. 1.

Kremsner, P. G. et al. Efficacy and safety of the CVnCoV SARS-CoV-2 mRNA vaccine candidate: results from Herald, a phase 2b/3, randomised, observer-blinded, placebo-controlled clinical trial in ten countries in Europe and Latin America. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(21\)00677-0](https://doi.org/10.1016/S1473-3099(21)00677-0) (2021).

2. 2.

Hoffmann, D. et al. CVnCoV and CV2CoV protect human ACE2 transgenic mice from ancestral B BavPat1 and emerging B.1.351 SARS-CoV-2. *Nat. Commun.* **12**, 4048 (2021).

3. 3.

Roth, N. et al. CV2CoV, an enhanced mRNA-based SARS-CoV-2 vaccine candidate, supports higher protein expression and improved immunogenicity in

rats. Preprint at <https://doi.org/10.1101/2021.05.13.443734> (2021).

4. 4.

Hoerr, I., Obst, R., Rammensee, H. G. & Jung, G. In vivo application of RNA leads to induction of specific cytotoxic T lymphocytes and antibodies. *Eur. J. Immunol.* **30**, 1–7 (2000).

5. 5.

Fotin-Mleczek, M. et al. Highly potent mRNA based cancer vaccines represent an attractive platform for combination therapies supporting an improved therapeutic effect. *J. Gene Med.* **14**, 428–439 (2012).

6. 6.

Fotin-Mleczek, M. et al. Messenger RNA-based vaccines with dual activity induce balanced TLR-7 dependent adaptive immune responses and provide antitumor activity. *J. Immunother.* **34**, 1–15 (2011).

7. 7.

Kubler, H. et al. Self-adjuvanted mRNA vaccination in advanced prostate cancer patients: a first-in-man phase I/IIa study. *J. Immunother. Cancer* **3**, 26 (2015).

8. 8.

Lutz, J. et al. Unmodified mRNA in LNPs constitutes a competitive technology for prophylactic vaccines. *NPJVaccines* **2**, 29 (2017).

9. 9.

Stitz, L. et al. A thermostable messenger RNA based vaccine against rabies. *PLoS Negl. Trop. Dis.* **11**, e0006108 (2017).

10. 10.

Schnee, M. et al. An mRNA vaccine encoding rabies virus glycoprotein induces protection against lethal infection in mice and correlates of protection in adult and newborn pigs. *PLoS Negl. Trop. Dis.* **10**, e0004746 (2016).

11. 11.

Petsch, B. et al. Protective efficacy of in vitro synthesized, specific mRNA vaccines against influenza A virus infection. *Nat. Biotechnol.* **30**, 1210–1216 (2012).

12. 12.

Aldrich, C. et al. Proof-of-concept of a low-dose unmodified mRNA-based rabies vaccine formulated with lipid nanoparticles in human volunteers: a phase 1 trial. *Vaccine* **39**, 1310–1318 (2021).

13. 13.

Pallesen, J. et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl Acad. Sci. USA* **114**, E7348–E7357 (2017).

14. 14.

Kirchdoerfer, R. N. et al. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.* **8**, 15701 (2018).

15. 15.

Yu, J. et al. DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science* **369**, 806–811 (2020).

16. 16.

Mercado, N. B. et al. Single-shot Ad26 vaccine protects against SARS-CoV-2 in rhesus macaques. *Nature* **586**, 583–588 (2020).

17. 17.

Chandrashekhar, A. et al. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science* **369**, 812–817 (2020).

18. 18.

Martinez, D. R. et al. Chimeric spike mRNA vaccines protect against Sarbecovirus challenge in mice. *Science* **373**, 991–998 (2021).

19. 19.

He, X. et al. Low-dose Ad26.COV2.S protection against SARS-CoV-2 challenge in rhesus macaques. *Cell* <https://doi.org/10.1016/j.cell.2021.05.040> (2021).

20. 20.

He, X. et al. Low-dose Ad26.COV2.S protection against SARS-CoV-2 challenge in rhesus macaques. *Cell* **184**, 3467–3473 (2021).

21. 21.

Dagotto, G. et al. Comparison of subgenomic and total RNA in SARS-CoV-2 challenged rhesus macaques. *J. Virol.* <https://doi.org/10.1128/JVI.02370-20> (2021).

22. 22.

McMahan, K. et al. Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature* **590**, 630–634 (2021).

23. 23.

Rauch, S. et al. mRNA-based SARS-CoV-2 vaccine candidate CVnCoV induces high levels of virus-neutralising antibodies and mediates protection in rodents. *NPJ Vaccines* **6**, 57 (2021).

24. 24.

Rauch, S. et al. mRNA vaccine CVnCoV protects non-human primates from SARS-CoV-2 challenge infection. Preprint at <https://doi.org/10.1101/2021.03.22.435960> (2020).

25. 25.

Polack, F. P. et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).

26. 26.

Baden, L. R. et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021).

27. 27.

Corbett, K. S. et al. Evaluation of the mRNA-1273 vaccine against SARS-CoV-2 in nonhuman primates. *N. Engl. J. Med.* **383**, 1544–1555 (2020).

28. 28.

Vogel, A. B. et al. BNT162b vaccines protect rhesus macaques from SARS-CoV-2. *Nature* **592**, 283–289 (2021).

29. 29.

Liu, C. et al. Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* <https://doi.org/10.1016/j.cell.2021.06.020> (2021).

30. 30.

Haas, E. J. et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet* **397**, 1819–1829 (2021).

31. 31.

Wu, K. et al. Serum neutralizing activity elicited by mRNA-1273 vaccine. *N. Engl. J. Med.* **384**, 1468–1470 (2021).

32. 32.

Wibmer, C. K. et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **27**, 622–625 (2021).

33. 33.

Wall, E. C. et al. Neutralising antibody activity against SARS-CoV-2 VOCs B.1.617.2 and B.1.351 by BNT162b2 vaccination. *Lancet* **397**, 2331–2333 (2021).

34. 34.

Philipp, M. & Santibanez, G. Preference of respiratory phases to perform reaction time tasks. *Act. Nerv. Super.* **30**, 153–155 (1988).

35. 35.

Feng, S. et al. Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nat. Med.* **27**, 2032–2040 (2021).

36. 36.

Lafon, E. et al. Potent SARS-CoV-2-specific T cell immunity and low anaphylatoxin levels correlate with mild disease progression in COVID-19 patients. *Front. Immunol.* **12**, 684014 (2021).

37. 37.

Schmidt, M. E. & Varga, S. M. The CD8 T cell response to respiratory virus infections. *Front. Immunol.* **9**, 678 (2018).

38. 38.

Abayasingam, A. et al. Long-term persistence of RBD⁺ memory B cells encoding neutralizing antibodies in SARS-CoV-2 infection. *Cell Rep. Med.* **2**, 100228 (2021).

39. 39.

Dan, J. M. et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, <https://doi.org/10.1126/science.abf4063> (2021).

40. 40.

Graham, B. S. Rapid COVID-19 vaccine development. *Science* **368**, 945–946 (2020).

Acknowledgements

We thank S. Gardner, G. Kennedy and R. Edmonston for their assistance and D. Maione and M.-T. Martin for their analysis of the manuscript. This work was supported by CureVac AG and the German Federal Ministry of Education and Research (BMBF; 01KI20703), the National Institutes of Health (CA260476), the Massachusetts Consortium on Pathogen Readiness and the Ragon Institute of MGH, MIT and Harvard. Development of CV2CoV is conducted in a collaboration of CureVac AG and GSK.

Author information

Author notes

1. These authors contributed equally: Makda S. Gebre, Susanne Rauch, Benjamin Petsch, Dan H. Barouch

Affiliations

1. Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

Makda S. Gebre, Jingyou Yu, Abishek Chandrashekhar, Noe B. Mercado, Xuan He, Jinyan Liu, Katherine McMahan, Victoria Giffin, David Hope, Shivani Patel, Daniel Sellers, Owen Sanborn, Julia Barrett & Dan H. Barouch

2. CureVac AG, Tübingen, Germany

Susanne Rauch, Nicole Roth, Stefan O. Mueller & Benjamin Petsch

3. Tufts University Cummings School of Veterinary Medicine, North Grafton, MA, USA

Amanda Martinot

4. University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

David R. Martinez & Ralph S. Baric

5. Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

Xiaowen Liu & Andrew C. Cole

6. Bioqual, Rockville, MD, USA

Laurent Pessant, Daniel Valentin, Zack Flinchbaugh, Jake Valley-Ogunro, Jeanne Muench, Renita Brown, Anthony Cook, Elyse Teow, Hanne Andersen & Mark G. Lewis

7. Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO, USA

Adrianus C. M. Boon

8. Ragon Institute of MGH, Ragon Institute of MGH, MIT and Harvard,
Cambridge, MA, USA

Dan H. Barouch

Contributions

S.R., B.P., S.O.M., N.R. and D.H.B. designed the study. M.S.G., J.Y., A. Chandrashekhar, N.B.M., X.H., J.L., K.M., A.M., D.R.M., R.S.B., A.C.M.B., V.G., D.H., S.P., D.S., O.S. and J.B. performed immunological and virological assays. X.L. and A.C.C. performed cytokine analysis. L.P., D.V., Z.F., J.Y.-O., J.M., R.B., A. Cook, E.T., H.A. and M.G.L. led the clinical care of the animals. M.S.G. and D.H.B. wrote the manuscript with all co-authors.

Corresponding authors

Correspondence to [Susanne Rauch](#) or [Dan H. Barouch](#).

Ethics declarations

Competing interests

S.R., B.P., N.R. and S.O.M. are employees of CureVac AG, Tübingen, Germany, a publicly listed company developing mRNA-based vaccines and immunotherapeutics. Authors may hold shares in the company. S.R., B.P. and N.R. are inventors on several patents on mRNA vaccination and use thereof. The other authors declare no competing interests.

Additional information

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

Extended data figures and tables

[Extended Data Fig. 1 Innate cytokine induction following mRNA immunization \(6/group\).](#)

Sera isolated 24h post first injection were analyzed for a panel of 19 cytokines associated with viral infection using a U-PLEX Viral Combo kit from Meso Scale

Discovery. Changes in cytokine levels above the detection limits were detectable for 9 cytokines. Each dot represents an individual animal, bars depict the median and the dotted line shows limit of detection. Statistical analysis was performed using two-tailed nonparametric Mann-Whitney test.

[Source data](#)

Extended Data Fig. 2 Neutralizing antibody titers against variants.

Animals (6/group) were vaccinated twice with 12 μ g of CVnCoV or CV2CoV on d0 and d28 or remained untreated as negative controls (sham). Sera isolated on d42 (week 6) were analyzed for pseudovirus neutralizing antibody titers against C.37 (Lambda), B.1.617.1 (Kappa) and B.1.617.2 (Delta) variants. Each dot represents an individual animal, bars depict the median and the dotted line shows limit of detection.

[Source data](#)

Extended Data Fig. 3 Memory B and T cell immune responses day 42 following immunization.

PBMCs from negative control (sham), CVnCoV or CV2CoV vaccinated animals (6/group) isolated on d42 of the experiment were stained for (a) RBD and (b) Spike-specific activated memory B cells and analyzed by high-parameter flow cytometry. IFN γ responses to pooled spike peptides were analyzed via ELISPOT (c). Each dot represents an individual animal, bars depict the median and the dotted line shows limit of detection. Statistical analysis was performed using two-tailed nonparametric Mann-Whitney test. PBMC = peripheral blood mononuclear cell; SFC = spot forming cells.

[Source data](#)

Extended Data Fig. 4 Binding and neutralizing antibody titers correlate with protection against SARS-CoV-2.

Summary of area under curve (AUC) viral load values following SARS-CoV-2 challenge in BAL and nasal swab samples (6/group) (a, b); antibody correlates of protection for binding antibodies (c, d) and neutralizing antibodies (e, f). Statistical analysis was performed using two-tailed nonparametric Mann-Whitney test. Correlations was analyzed by two-sided Spearman rank-correlation test. NAbs = neutralizing antibodies, BAL = bronchoalveolar lavage NS = nasal swab.

[Source data](#)

Extended Data Fig. 5 Infectious virus titers after SARS-CoV-2 challenge (6/group).

Infectious virus titers of BAL and nasal swab samples collected 2 days post challenge were analyzed by TCID₅₀ assays. Each dot represents an individual animal, bars depict the median and the dotted line shows limit of detection. Statistical analysis was performed using two-tailed nonparametric Mann-Whitney test.

[Source data](#)

Extended Data Fig. 6 Post-challenge binding and neutralizing antibody responses (6/group).

Negative control (sham) or animals vaccinated on d0 and d28 of the experiment with 12 µg of CVnCoV or CV2CoV as indicated were subjected to challenge infection using 1.0×10^5 TCID₅₀ SARS-CoV-2 via intranasal (IN) and intratracheal (IT) routes.

(a) Titers of RBD binding antibodies and (b) pseudovirus neutralizing antibodies against ancestral SARS-CoV-2 strain were evaluated before (week 8) and a week after challenge infection (week 9). Each dot represents an individual animal, bars depict the median and the dotted line shows limit of detection. Statistical analysis was performed using two-tailed nonparametric Mann-Whitney test. NAbs = neutralizing antibodies.

[Source data](#)

Extended Data Fig. 7 CVnCoV and CV2CoV protect the lungs from pathological changes upon viral challenge (6/group).

Eight lung lobes (4 sections from right and left, caudal to cranial) were assessed and scored (1-4) for each of the following lesions: 1) Interstitial inflammation and septal thickening 2) Eosinophilic interstitial infiltrate 3) Neutrophilic interstitial infiltrate 4) Hyaline membranes 5) Interstitial fibrosis 6) Alveolar infiltrate, macrophage 7) Alveolar/Bronchoalveolar infiltrate, neutrophils 8) Syncytial cells 9) Type II pneumocyte hyperplasia 10) Broncholar infiltrate, macrophage 11) Broncholar infiltrate, neutrophils 12) BALT hyperplasia 13) Bronchiolar/peribronchiolar inflammation 14) Perivascular, mononuclear infiltrates 15) Vessels, endothelialitis. Each feature assessed was assigned a score of 0 = no significant findings; 1 = minimal; 2 = mild; 3 = moderate; 4 = marked/severe. (a) Cumulative scores per animal (b) Cumulative scores per lung lobe. Individual animals are represented by symbols. Representative histopathology from sham vaccinated (c-h), CnVCoV vaccinated (i, j), and Cv2CoV vaccinated (k, l) animals showing (c, d, inset) alveolar macrophage infiltrate, (e, f, inset) syncytial cells (arrowheads) and type II pneumocyte hyperplasia,

inset (g, h, inset) bronchiolar epithelial necrosis with neutrophilic infiltrates (**i**) alveolar neutrophilic infiltrate and alveolar septal thickening (**j**) focal consolidation with inflammation composed of macrophages, neutrophils, and syncytial cells (**k**) focal pneumocyte hyperplasia, syncytial cells and inflammatory infiltrates (**l**) peribronchiolar inflammation. Statistical analysis was performed using two-tailed nonparametric Mann-Whitney test. Scale bars: 100 microns (**c**), 50 microns (**e, g**) 20 microns (**i-l**). BALT bronchus associated lymphoid tissue.

[Source data](#)

Supplementary information

[Reporting Summary](#)

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](#)

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

Gebre, M.S., Rauch, S., Roth, N. *et al.* Optimization of non-coding regions for a non-modified mRNA COVID-19 vaccine. *Nature* **601**, 410–414 (2022).
<https://doi.org/10.1038/s41586-021-04231-6>

- Received: 13 August 2021
- Accepted: 11 November 2021
- Published: 18 November 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04231-6>

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/s41586-021-04231-6>

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

- Article
- [Published: 05 January 2022](#)

Behavioural immune landscapes of inflammation

- [Georgiana Crainiciuc](#)^{1,na1},
- [Miguel Palomino-Segura](#) ORCID: orcid.org/0000-0003-1614-1222^{1,na1},
- [Miguel Molina-Moreno](#)²,
- [Jon Sicilia](#)^{1,3},
- [David G. Aragones](#) ORCID: orcid.org/0000-0003-4035-0972⁴,
- [Jackson Liang Yao Li](#)^{1,5},
- [Rodrigo Madurga](#) ORCID: orcid.org/0000-0002-1381-2599⁶,
- [José M. Adrover](#) ORCID: orcid.org/0000-0002-1395-5477¹,
- [Alejandra Aroca-Crevillén](#) ORCID: orcid.org/0000-0002-2481-0863¹,
- [Sandra Martin-Salamanca](#)¹,
- [Alfonso Serrano del Valle](#) ORCID: orcid.org/0000-0001-9343-1050¹,
- [Sandra D. Castillo](#) ORCID: orcid.org/0000-0002-7007-3155^{7,8},
- [Heidi C. E. Welch](#) ORCID: orcid.org/0000-0001-7865-7000⁹,
- [Oliver Soehnlein](#) ORCID: orcid.org/0000-0002-7854-0694¹⁰,
- [Mariona Graupera](#) ORCID: orcid.org/0000-0003-4608-4185^{7,8},
- [Fátima Sánchez-Cabo](#)³,
- [Alexander Zarbock](#)¹¹,
- [Thomas E. Smithgall](#) ORCID: orcid.org/0000-0001-5238-3806¹²,
- [Mauro Di Pilato](#)^{13,14},
- [Thorsten R. Mempel](#) ORCID: orcid.org/0000-0001-8635-5424¹³,
- [Pierre-Louis Tharaux](#) ORCID: orcid.org/0000-0002-6062-5905¹⁵,
- [Santiago F. González](#) ORCID: orcid.org/0000-0003-4166-7664¹⁶,
- [Angel Ayuso-Sacido](#)^{6,17},
- [Lai Guan Ng](#) ORCID: orcid.org/0000-0003-1905-3586⁵,

- [Gabriel F. Calvo⁴](#),
- [Iván González-Díaz²](#),
- [Fernando Díaz-de-María](#) ORCID: [orcid.org/0000-0002-6437-914X²](https://orcid.org/0000-0002-6437-914X)
&
- [Andrés Hidalgo](#) ORCID: [orcid.org/0000-0001-5513-555X^{1,18}](https://orcid.org/0000-0001-5513-555X)

[Nature](#) volume **601**, pages 415–421 (2022)

- 18k Accesses
- 435 Altmetric
- [Metrics details](#)

Subjects

- [Cellular imaging](#)
- [Imaging the immune system](#)

Abstract

Transcriptional and proteomic profiling of individual cells have revolutionized interpretation of biological phenomena by providing cellular landscapes of healthy and diseased tissues^{1,2}. These approaches, however, do not describe dynamic scenarios in which cells continuously change their biochemical properties and downstream ‘behavioural’ outputs^{3,4,5}. Here we used 4D live imaging to record tens to hundreds of morpho-kinetic parameters describing the dynamics of individual leukocytes at sites of active inflammation. By analysing more than 100,000 reconstructions of cell shapes and tracks over time, we obtained behavioural descriptors of individual cells and used these high-dimensional datasets to build behavioural landscapes. These landscapes recognized leukocyte identities in the inflamed skin and trachea, and uncovered a continuum of neutrophil states inside blood vessels, including a large, sessile state that was embraced by the underlying endothelium and associated with pathogenic inflammation. Behavioural screening in 24 mouse mutants identified the

kinase Fgr as a driver of this pathogenic state, and interference with Fgr protected mice from inflammatory injury. Thus, behavioural landscapes report distinct properties of dynamic environments at high cellular resolution.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Behavioural maps capture immune identities.



Fig. 2: Behavioural landscape of intravascular inflammation.

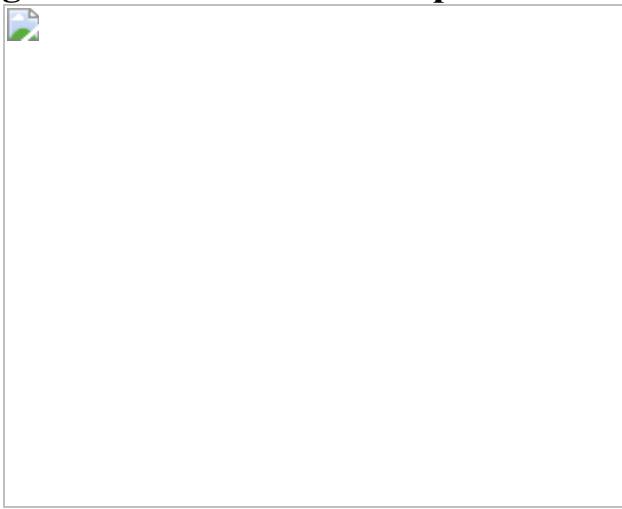


Fig. 3: Screening for drivers of pathogenic intravascular behaviours.

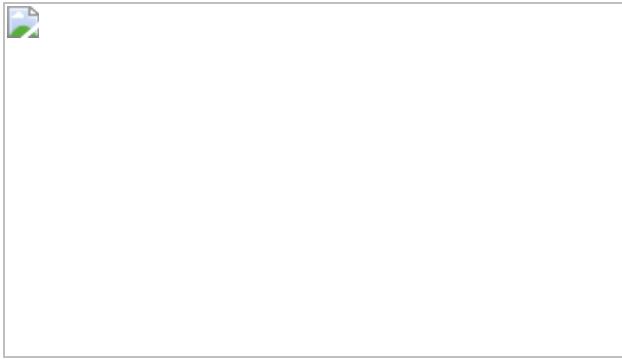
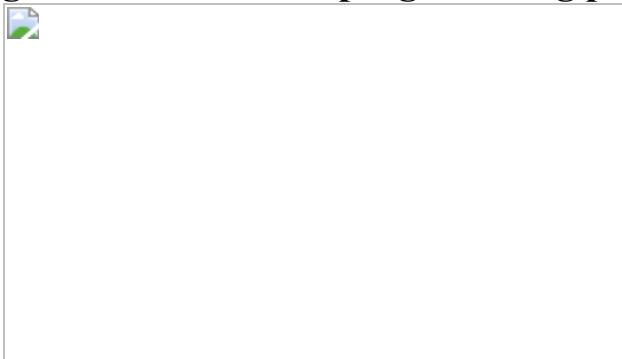


Fig. 4: Behavioural reprogramming protects from inflammation.



Data availability

All data and materials used in the study are available to any researcher for purposes of reproducing or extending these analyses. [Source data](#) are provided with this paper.

Code availability

All code used are available to any researcher for purposes of reproducing or extending these analyses. The code for ACME is available at <https://doi.org/10.5281/zenodo.5638537>.

References

1. 1.

Kwok, I. et al. Combinatorial single-cell analyses of granulocyte–monocyte progenitor heterogeneity reveals an early uni-potent

neutrophil progenitor. *Immunity* **53**, 303–318.e5 (2020).

2. 2.

Amit, I., Regev, A. & Hacohen, N. Strategies to discover regulatory circuits of the mammalian immune system. *Nat. Rev. Immunol.* **11**, 873–880 (2011).

3. 3.

Lämmermann, T. et al. Neutrophil swarms require LTB4 and integrins at sites of cell death in vivo. *Nature* **498**, 371–375 (2013).

4. 4.

Sreeramkumar, V. et al. Neutrophils scan for activated platelets to initiate inflammation. *Science* **346**, 1234–1238 (2014).

5. 5.

Woodfin, A. et al. The junctional adhesion molecule JAM-C regulates polarized transendothelial migration of neutrophils in vivo. *Nat. Immunol.* **12**, 761–769 (2011).

6. 6.

Papalexis, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* **18**, 35–45 (2018).

7. 7.

Sumen, C., Mempel, T. R., Bazo, I. B. & von Andrian, U. H. Intravital microscopy: visualizing immunity in context. *Immunity* **21**, 315–329 (2004).

8. 8.

Lelkes, E., Headley, M. B., Thornton, E. E., Looney, M. R. & Krummel, M. F. The spatiotemporal cellular dynamics of lung

immunity. *Trends Immunol.* **35**, 379–386 (2014).

9. 9.

Yamada, K. M. & Sixt, M. Mechanisms of 3D cell migration. *Nat. Rev. Mol. Cell Biol.* **20**, 738–752 (2019).

10. 10.

Goh, C. C. et al. Real-time imaging of dendritic cell responses to sterile tissue injury. *J. Invest. Dermatol.* **135**, 1181–1184 (2015).

11. 11.

Goh, C. C. et al. The impact of ischemia–reperfusion injuries on skin resident murine dendritic cells. *Eur. J. Immunol.* **48**, 1014–1019 (2018).

12. 12.

Adrover, J. M. et al. Programmed ‘disarming’ of the neutrophil proteome reduces the magnitude of inflammation. *Nat. Immunol.* **21**, 135–144 (2020).

13. 13.

Adrover, J. M. et al. A neutrophil timer coordinates immune defense and vascular protection. *Immunity* **50**, 390–402.e10 (2019).

14. 14.

Hasenberg, A. et al. Catchup: a mouse model for imaging-based tracking and modulation of neutrophil granulocytes. *Nat. Methods* **12**, 445–452 (2015).

15. 15.

Hidalgo, A. et al. Heterotypic interactions enabled by polarized neutrophil microdomains mediate thromboinflammatory injury. *Nat.*

Med. **15**, 384–391 (2009).

16. 16.

Mócsai, A., Walzog, B. & Lowell, C. A. Intracellular signalling during neutrophil recruitment. *Cardiovasc. Res.* **107**, 373–385 (2015).

17. 17.

Weir, M. C. et al. Selective inhibition of the myeloid Src-family kinase Fgr potently suppresses AML cell growth in vitro and in vivo. *ACS Chem. Biol.* **13**, 1551–1559 (2018).

18. 18.

García-Prieto, J. et al. Neutrophil stunning by metoprolol reduces infarct size. *Nat. Commun.* **8**, 14780 (2017).

19. 19.

Hirahashi, J. et al. Mac-1 (CD11b/CD18) links inflammation and thrombosis after glomerular injury. *Circulation* **120**, 1255–1265 (2009).

20. 20.

McArdle, S. et al. Migratory and dancing macrophage subsets in atherosclerotic lesions. *Circ. Res.* **125**, 1038–1051 (2019).

21. 21.

Mempel, T. R., Henrickson, S. E. & von Andrian, U. H. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* **427**, 154–159 (2004).

22. 22.

Boulch, M. et al. A cross-talk between CAR T cell subsets and the tumor microenvironment is essential for sustained cytotoxic activity.

Sci. Immunol. **6**, eabd4344 (2021).

23. 23.

Ivanovitch, K., Temiño, S. & Torres, M. Live imaging of heart tube development in mouse reveals alternating phases of cardiac differentiation and morphogenesis. *eLife* **6**, e30668 (2017).

24. 24.

Lindquist, R. L. et al. Visualizing dendritic cell networks in vivo. *Nat. Immunol.* **5**, 1243–1250 (2004).

25. 25.

Faust, N., Varas, F., Kelly, L. M., Heck, S. & Graf, T. Insertion of enhanced green fluorescent protein into the lysozyme gene creates mice with green fluorescent granulocytes and macrophages. *Blood* **96**, 719–726 (2000).

26. 26.

Palomino-Segura, M., Virgilio, T., Morone, D., Pizzagalli, D. U. & Gonzalez, S. F. Imaging cell interaction in tracheal mucosa during influenza virus infection using two-photon intravital microscopy. *J. Vis. Exp.* 58355 (2018).

27. 27.

Bauer, C. A. et al. Dynamic T_{reg} interactions with intratumoral APCs promote local CTL dysfunction. *J. Clin. Invest.* **124**, 2425–2440 (2014).

28. 28.

Tran Cao, H. S. et al. Development of the transgenic cyan fluorescent protein (CFP)-expressing nude mouse for ‘technicolor’ cancer imaging. *J. Cell. Biochem.* **107**, 328–334 (2009).

29. 29.

Devi, S. et al. Neutrophil mobilization via plerixafor mediated CXCR4 inhibition arises from lung demargination and blockade of neutrophil homing to the bone marrow. *J. Exp. Med.* **210**, 2321–2336 (2013).

30. 30.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. In *MICCAI 2016. Lecture Notes in Computer Science* Vol. 9901 (eds Ourselin, S. et al.) (Springer, 2016).

31. 31.

Welch, G. & Bishop, G. An introduction to the Kalman Filter. *In Pract.* (2006).

32. 32.

Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

33. 33.

Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* 4th edn (Springer, 2002).

34. 34.

Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

35. 35.

Lloyd, S. P. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982).

36. 36.

Henique, C. et al. Genetic and pharmacological inhibition of microRNA-92a maintains podocyte cell cycle quiescence and limits crescentic glomerulonephritis. *Nat. Commun.* **8**, 1829 (2017).

Acknowledgements

We thank all members of the Hidalgo laboratory and M. Desco for discussion; P. Frenette for inspiring this study; C. C. Goh and E. Y. Kim for seeding imaging experiments; the electron microscopy unit from the faculty of Medicine of Universidad Autonoma de Madrid for help with experiments; E. Marín, L. Cabezuela, E. Santos, R. Mota and the animal facility at CNIC for animal husbandry, animal procedures and histology; J. Rossaint, M. Gunzer, J.A. Enriquez, A. Mocsai, R.W. Hendricks, G. Sabio, M. Sperandio, E. Hirsch and B. Walzog for the generous gift of mutant mice; and C. Torroja, D. Jiménez and M. Desco for technical advice. This study was supported by RTI2018-095497-B-I00 from Ministerio de Ciencia e Innovación (MCIN), HR17_00527 from Fundación La Caixa, Transatlantic Network of Excellence (TNE-18CVD04) from the Leducq Foundation, and FET-OPEN (no. 861878) from the European Commission to A.H. M.P-S. is supported by a Federation of European Biochemical Societies and the EMBO ALTF (no. 1142-2020) long-term fellowships. J.S. is supported by a fellowship (PRE2019-089130) from MICINN and A.A.-C. is supported by fellowship CF/BQ/DR19/11740022 from La Caixa Foundation. J.L.Y.L. was supported by A*STAR and a Juan de la Cierva JCI-2017-33136 Fellowship from MICINN. S.D.C. is a recipient of a Marie Skłodowska-Curie fellowship (749731). M.G. is supported by SAF2017-89116R-P from MCIN and HR18_00120 from la Fundación La Caixa. T.R.M. is supported by grant NIH AI163223, L.G.N. is supported by SIgN core funding from A*STAR, and G.F.C. is supported by MCIN/AEI/10.13039/501100011033 (grant PID2019-110895RB-I00) and by Junta de Comunidades de Castilla-La Mancha (SBPLY/19/180501/000211). F.S.-C. is supported by MCIN (grant RTI2018-102084-B-I00), O.S. is supported by the Leducq Foundation (TNE-18CVD04), F.D.-d.-M. is supported by MCIN (TEC2017-84395-P), and T.E.S. is supported by the National Cancer Institute, NIH grant CA233576. The CNIC is supported by the MCIN and the Pro-CNIC Foundation.

Author information

Author notes

1. These authors contributed equally: Georgiana Crainiciuc, Miguel Palomino-Segura

Affiliations

1. Area of Cell and Developmental Biology, Centro Nacional de Investigaciones Cardiovasculares Carlos III, Madrid, Spain

Georgiana Crainiciuc, Miguel Palomino-Segura, Jon Sicilia, Jackson Liang Yao Li, José M. Adrover, Alejandra Aroca-Crevillén, Sandra Martín-Salamanca, Alfonso Serrano del Valle & Andrés Hidalgo

2. Department of Signal Processing and Communication, Universidad Carlos III de Madrid, Madrid, Spain

Miguel Molina-Moreno, Iván González-Díaz & Fernando Díaz-de-María

3. Bioinformatics Unit, Centro Nacional de Investigaciones Cardiovasculares Carlos III, Madrid, Spain

Jon Sicilia & Fátima Sánchez-Cabo

4. Department of Mathematics & MOLAB-Mathematical Oncology Laboratory, Universidad de Castilla-La Mancha, Ciudad Real, Spain

David G. Aragones & Gabriel F. Calvo

5. Singapore Immunology Network (SIN), A*STAR, Biopolis, Singapore

Jackson Liang Yao Li & Lai Guan Ng

6. Faculty of Experimental Sciences and Faculty of Medicine,
Universidad Francisco de Vitoria, Madrid, Spain

Rodrigo Madurga & Angel Ayuso-Sacido

7. Endothelial Pathobiology and Microenvironment Group, Josep Carreras Leukaemia Research Institute (IJC), 08916 Badalona, Barcelona, Spain

Sandra D. Castillo & Mariona Graupera

8. Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Instituto de Salud Carlos III, Madrid, Spain

Sandra D. Castillo & Mariona Graupera

9. Signalling Programme, Babraham Institute, Cambridge, UK

Heidi C. E. Welch

10. Institute for Experimental Pathology, Center for Molecular Biology of Inflammation, Westfälische Wilhelms-Universität, Münster, Germany

Oliver Soehnlein

11. Department of Anesthesiology, Intensive Care and Pain Medicine, University Hospital Münster, Münster, Germany

Alexander Zarbock

12. Department of Microbiology and Molecular Genetics, University of Pittsburgh, School of Medicine, Pittsburgh, PA, USA

Thomas E. Smithgall

13. Center for Immunology and Inflammatory Diseases at Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

Mauro Di Pilato & Thorsten R. Mempel

14. Department of Immunology, the University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Mauro Di Pilato

15. Université de Paris, Paris Cardiovascular Center, Inserm, Paris, France

Pierre-Louis Tharaux

16. Institute for Research in Biomedicine, Università della Svizzera Italiana, Bellinzona, Switzerland

Santiago F. González

17. Brain Tumor Laboratory, Fundación Vithas, Grupo Hospitales Vithas, Madrid, Spain

Angel Ayuso-Sacido

18. Vascular Biology and Therapeutics Program and Department of Immunobiology, Yale University School of Medicine, New Haven, USA

Andrés Hidalgo

Contributions

Conceptualization: A.H. Methodology: M.P.-S., F.S.-C., A.Z., T.E.S., P.-L.T., H.C.E.W., O.S., M.M.-M., G.F.C., I.G.-D., F.D.-d.-M. and A.H. Investigation: G.C., M.P.-S., M.M.-M., J.S., D.G.A., G.F.C., J.L.Y.L., R.M., J.M.A., A.A.-C., S.M.-S., A.S.d.V., S.D.C. and M.D.P. Visualization and intravital microscopy: G.C., M.P.-S., M.M.-M., J.S. and D.G.A. Image data analysis: M.P.-S., I.G.-D., M.M.-M., J.L.Y.L. and M.D.P. Myocardial infarction and glomerulonephritis experiments: G.C. and S.M.-S. Funding acquisition: A.H., F.S.-C., P.-L.T., L.G.N. and G.F.C. Supervision: P.-L.T., S.F.G., T.R.M., A.A.-S., L.G.N., G.F.C., I.G.-D., F.D.-d.-M. and A.H.

Writing, original draft: M.P.-S. and A.H. Writing, review and editing: All authors. Contribution note: M.M.-M. and J.S. contributed equally as secondary authors in this paper.

Corresponding author

Correspondence to [Andrés Hidalgo](#).

Ethics declarations

Competing interests

A.H. is a consultant for Flagship Pioneering. The other authors declare no competing interests.

Peer review information

Nature thanks Klaus Ley 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 Selection of parameters for behavioral analyses.](#)

(A) Application of size filters. Left, representative images showing use of >40 voxel filter to eliminate subcellular objects in the trachea experiments. Comparison of the raw vs. surface reconstructed objects (see merged) eliminates fragment-like objects, as shown in the insets. Right, tSNE representation of the trachea dataset in which the filter for cell size was set

to 0 (no filter) or 160 voxels, showing objects with sizes below 40 voxels (threshold used in our experiments). The number of objects for each representation are shown in brackets. Note that segregation of neutrophils and DCs into different visual clusters is compromised in the absence of filter. **(B)** Workflow for parameter selection. 4D images were analyzed to extract morphometric and kinetic parameters (118 in our experiments using Imaris software). We performed supervised selection of the best characteristics following criteria of redundancy, biological features of interest in the specific biological setup, or removal of non-biological parameters such as arbitrary position. In parallel we generated correlation networks for all parameters and each experiment (118 in the experiments reported here), and we visualized the distribution of the selected parameters in the correlation networks (see below). Finally, we reduced dimensionality using the selected parameters to identify behavioral clusters for further validation. For the “training” experiments shown here, where cell identities were known, we determined LRI/ARI to complement our correlation networks with the power of each parameter to classify cells correctly. The workflow is fully adaptable to other image analysis tools, as well as algorithms to establish correlation between parameters and for dimensional reduction, including elastic net regression methods. **(C)** Correlation networks used for the experiments shown in Fig. 1 with Imaris image analysis. Networks on the left column highlight the specific 31 parameters selected, which are identified by the number code shown in Supplementary Table 2. Correlation networks on the right column correspond to the three datasets shown in Fig. 1 (influenza infection in the trachea, ischemia-reperfusion in skin and laser injury in skin), showing parameters as circles whose diameters are proportional to their LRI, as well as the positive (red) and negative (blue) correlations between each pair of parameters. The thickness of the links is proportional to the absolute value of the Pearson correlation coefficient for each pair, and the distance between parameters reflect the similarity of the Pearson coefficients with the rest of parameters. **(D)** Violin plots showing LRI/ARI values for all 118 vs. the selected 31 parameters. Lines represent medians. Data compared by Mann-Whitney non-parametric test. **(E)** Heatmap showing the LRI/ARI values for each of the 31 selected parameters for each experiment, as well as the geometric mean for the three experiments combined, reflecting the average power of each parameter in our experiments. **(F)** Quality of parameter selection. Top,

number of parameters selected, and predictive power (LRI or ARI in red) of the parameters selected by Lasso regression compared with our list of 31 selected parameters. Bottom, comparison of the distribution of the 25 parameters selected by Lasso regression and our 31 selected parameters across the correlation network for the trachea experiment. **(G)** tSNE plots generated by considering only the morphometric, or only the kinetic parameters, or both combined. Donut plots show the distribution of the analyzed cell types (macrophages, DCs and neutrophils) in each cluster. Note that the accuracy in identifying specific cell types for each cluster is always highest when both classes of parameters are combined. **(H)** tSNE plots generated by considering all 118 parameters or only the selected 31 across all three experimental setups (influenza infection in the trachea, ischemia-reperfusion in skin and laser injury in skin). Cell classification per cluster was better for the selected 31 parameters. **(I)** tSNE plots showing the classification of cells into clusters by using all 118 parameters and a standard single cell analytical pipeline (Seurat_v4). Donut plots indicate the distribution of the analyzed cell types (macrophages, DCs and neutrophils) in each cluster.

[Source data](#)

[Extended Data Fig. 2 Behavioral landscape of the infected trachea.](#)

(A) Heatmap of the 31 behavioral parameters used for the trachea infection analysis. For a full list of all extracted parameters please refer to Supplementary Table [2](#). **(B)** Pearson correlation matrix for all 118 parameters extracted from the trachea imaging experiment, with the selected parameters marked in red font. **(C)** Segmentation of cells from the trachea by four different combinations of morpho-kinetic parameters. We randomly chose 1, 6, 15, 25 and 31 parameters (list of parameter code numbers shown at right) and used them to represent the separation of neutrophils and DCs using tSNE. The original set of parameters used in Fig. [1c](#) is at the top. Parameters are ordered from higher to lower LRI (left to right) to better visualize the classification value of each parameter used in the plot analyses. **(D)** LRI (score of cell identities) are proportional to the number of parameters extracted from the imaging experiments and

combined to infer identities. Violin plots show the distribution of LRIs when using 1-5 parameters to classify cells in the trachea experiment, assuming that only sets of 5, 25, 31, 50 or 118 parameters are available for analysis. Note that the LRIs shown here are for the full 118 parameter set, and are not comparable with the 31 subset of selected parameters, which feature higher LRI values, as shown in the violin plots on the far right. **(E)** Individual analyses of the behavior of DCs and neutrophils from the original dataset, shown as tSNE plots for each population. Each behavioral parameter can be visualized and compared across cell subsets and parameters to infer positive or negative correlations, as shown for *Distance to DC* which negatively correlates with cell speeds in the Pearson correlation matrix of the 31 parameters used in the final analysis **(F)**.

[Source data](#)

Extended Data Fig. 3 Behavioral landscape of the skin under ischemia-reperfusion.

(A) Representative image of I/R injury of the skin (original image on top; reconstruction of volumes and tracks at bottom). Below, examples of a typical GFP^{lo} macrophage, a GFP^{hi} neutrophil and an YFP+ DC used to classify the cells post-analysis. **(B)** Heatmap of all parameters and classified by cluster (0, 1 y 2) from the plot in Fig. [1h](#), and further divided into subclusters shown in **(C)**. Below, expression plots of selected parameters. **(C)** tSNE plot showing all subclusters identified in the heatmap in **(B)**. Donut plots indicate the fraction of neutrophils, DCs and macrophages in each cluster. Bottom panels show the behavioral maps generated by back-gating each cluster into the original position for each cell so that maps show the position of cells with the same behavioral profile.

Extended Data Fig. 4 Behavioral landscape of laser injury in the skin.

(A) Representative image of laser burn injury (original image left; reconstruction of volumes and tracks at right), **(B)** Heatmap of the all scored parameters, showing DCs and neutrophils. Expression tSNE plots of selected parameters are shown at bottom. **(C)** Individual analyses of the

behavior of DCs and neutrophils from the original dataset, shown as tSNE plots for each population. Each behavioral parameter can be visualized and compared across cell subsets and parameters to infer random or gradient distribution for each population. For example, the location of the laser injury can be extracted as a parameter (left, yellow arrowhead) that shows graded behaviors of neutrophils relative to their distance to the wound, but not for DCs. **(D)** Examples of behavioral maps generated by projecting the intensity of specific parameters onto the XY location of individual cells at all time points. Actual image, plot-map by cell type and behavioral maps are shown. **(E)** Sub-clustering identifies two behavioral clusters of neutrophils and one for DCs (top), which were projected back onto their corresponding xyz position thus giving a profile of the distribution of behavioral clusters in the skin anatomy (middle). The neutrophil clusters feature differences in various parameters, as shown in the expression plots (arrowheads in the bottom tSNE plots). **(F)** Representative image of regulatory T cells (Treg) and cytotoxic T cells (CTL) in a CT26 carcinoma (red outline) in the skin, and tSNE plots of the cells classified by behavioral phenotype and by cell type. Donut plots show the match between both classifications. **(G)** Heatmap of the differentially scored parameters discriminating CTLs and Tregs. **(H–J)** Behavioral landscapes and maps of CTLs in carcinoma-bearing mice **(H)**, neutrophils inside or outside inflamed vessels **(I)**, and bone marrow neutrophils before and after administration of the mobilizing chemokine CXCL1 **(J)**. Donut plots and expression plots illustrate the correlation between behavioral patterns or parameters and their localization in tissues. Dashed lines in the behavioral maps in **(H–I)** delineate tumor-stroma or vessel-parenchyma borders, respectively. Data are from one experiment per condition to visualize the distribution of cells in a single anatomical area.

[Source data](#)

Extended Data Fig. 5 Neutrophil states inside inflamed venules.

(A) Analysis of the cremaster dataset using Imaris software and UMAP representation show less defined behavioral clusters than using ACME (compare with Fig. [2d–f](#)). Donut plots show the distribution of clusters in control and platelet-depleted mice. **(B)** Anomalous morphometric

reconstructions of fast rolling cells, shown in top and side 3D views of cells moving at different speeds. Firmly adherent B1 neutrophils are shown for reference. **(C)** Rapid changes in morphology for neutrophils in the B2 group, following inchworm-type crawling during a 90 s recording; scale bar, 10 μ m. **(D)** Membrane extensions (yellow arrowheads) forming around large oblate neutrophils in the B3 group, but not from B1 or B2; scale bar, 5 μ m. **(E)** Representative micrograph of an inflamed vessel from Ly6G^{Cre}; Rosa26^{tdTom} mouse with several neutrophils exhibiting B2 and B3 behavioral profiles (arrowheads), and “footprints” beneath B3 cells; scale bar, 10 μ m. The presence of the footprints for each behavior is quantified in **(F)**, where n is number of neutrophils analyzed. **(G)** Micrographs and quantification of CD11b expression measured by in vivo imaging across the different behavioral groups, with rolling neutrophils included as reference cells; data is from the indicated number of cells (in brackets), from 6 mice per group. **(H)** Micrographs and quantification of the number of beads phagocytosed by neutrophils from each behavioral group, including rolling cells; n is the number of cells (in brackets) from 6 mice analyzed per group. Scale bar, 5 μ m. **(I)** Representative 3D image of an inflamed cremaster vessel showing examples of B2 and B3 neutrophils (left image), which were examined for extravasation across the endothelial wall over time (arrowheads in insets, right). **(J)** Percent of B3-type neutrophils that localize in junctional vs. non-junctional areas, and **(K)** the frequency of transendothelial migration (TEM) for each behavioral group of neutrophils; n is 5 mice per group, with the indicated number of analyzed cells (brackets). All bar graphs show mean \pm SEM and data were analyzed by one-way ANOVA with Tukey’s multigroup comparison test (**H, K**) or unpaired two-tailed t-test (**J**). Number of analyzed cells per group from 3-5 mice each are indicated in brackets.

[Source data](#)

[Extended Data Fig. 6 Transitional states of neutrophils in vessels.](#)

(A) Heatmap of all parameters across all behaviors, including the three subgroups in B2. **(B)** UMAP based on hierarchical clustering to identify two additional behavioral clusters within B2. **(C)** Distribution of cells in each

sub-cluster B2.1, B2.2 and B2.3 for the indicated parameters, showing for example that cells B2.3 feature sizes and distances to the vessel wall similar to those of B3. Data analyzed by one-way ANOVA. **(D)** Transitions between behavioral clusters shown graphically in the UMAP (left) and quantified at right. **(E)** Scheme illustrating the most common transitions typically involving passage through B2, suggesting that this is an obligate transitional stage for neutrophils in inflamed vessels. Drawings in each group represent the silhouettes of representative cells at different times as in Fig. 2h.

[Source data](#)

Extended Data Fig. 7 Track parameters in the behavioral screening.

(A) Heatmap of the differentially scored behaviors among the three main behavioral groups (B1, B2 and B3). Outlined in red are the specific behaviors chosen for our screening in Fig. 3. Note that “tortuosity” is an inverse measure of “directionality”, which was used in our screening. **(B)** Speed and directionality obtained by epifluorescence (2D) analysis of cremasteric venules in mice with mixed chimeric bone marrow of wild-type^{DsRed} and non-fluorescent mutant donors, which provided internal controls for each group. Thick lines show means; The number of analyzed cells per group is shown in brackets as (control, mutant), and were obtained from at least 3 mice per group. Data analyzed by unpaired two-tailed t-test.

[Source data](#)

Extended Data Fig. 8 Morphometric parameters in the behavioral screening.

Ellipticity prolate and H/L ratios measured for individual cells in static 3D reconstructions from 24 mutant and 3 control groups as summarized in Fig. 3. Values are from cremasteric venules in mice with mixed chimeric bone marrow of wild-type^{DsRed} and non-fluorescent mutant donors, which provided internal controls for each group. Thick lines show means; The number of analyzed cells per group is shown in brackets as (control,

mutant), and were obtained from at least 3 mice per group. Data analyzed by unpaired two-tailed t-test analysis.

[Source data](#)

Extended Data Fig. 9 Protection from myocardial injury by targeting Fgr.

(A) Micrographs of NETs (positive for citH3 and MPO; red) and vessels (blue) in cremasteric venules of wild-type subjected or not to I/R, and *Fgr*^{-/-} mice subjected to I/R. Right, quantification of NETs per tissue volume; Data shown as mean ± SEM and n are number of mice per group. (B) Competitive recruitment of wild-type and *Fgr*^{-/-} neutrophils to the peritoneal cavity after zymosan injection, or to the bronchoalveolar space of lungs after LPS instillation in mixed chimeric mice; n are numbers of mice analyzed. *Selplg*^{-/-} neutrophils are shown for comparison of impaired migration. Values are normalized to reference wild-type competitors across the different groups and given as migration efficiencies. Data shown as mean ± SEM and n are number of mice analyzed. (C) Micrographs of Weibel-Palade bodies (WPB) and vacuoles in myocardial vessels after sham or ischemic challenge, which are quantified in (D). These measures of vascular damage are dependent on neutrophils, as shown after experimental depletion with 1A8 antibody (E). Data shown as mean ± SEM, and n is the number of micrographs analyzed, from 2 mice. (F) Effect of the Fgr antagonist TL02-59 in myocardial death upon ischemia-reperfusion, when given after ischemia at the time of reperfusion. Micrographs of heart sections at left illustrate the protective effect on myocardial death (outlined whitish regions). Data are normalized to the area at risk (AAR) and shown as mean from 4 mice per group. (G) Combination of neutrophil depletion with 1A8 antibody, and Fgr deficiency in transplanted mice. The infarcted areas are normalized with the areas at risk; n is number of mice per group. (H) Combination of neutrophil depletion with the Fgr agonist TL02-59. Data shown as mean ± SEM; n are mice per group. (I) Combination of the Fgr inhibitor in hematopoietic *Fgr*^{-/-} mice, with no effect in further protecting from myocardial death; n are mice per group. (J) Myocardial fibrosis (left ventricle) determined by hematoxylin and eosin staining in

control wild-type and *Fgr*^{-/-} mice subjected to permanent ischemia and analyzed after 28 days. The fibrosis area is represented at right; n are mice per group. All data from (A, B) was analyzed by one-way ANOVA with Tukey's multiple comparison test; (G–I) was analyzed by two-way ANOVA with Tukey's multiple comparisons test. All other panels are compared by two-tailed unpaired-t test (C–F and J). The number of replicates (n) per group is indicated in each panel.

[Source data](#)

Extended Data Fig. 10 Protection from nephrotoxic injury by targeting Fgr.

(A) Schematic of the nephrotoxic injury model (top) and setup of conditions combining endotoxin (LPS) with increasing amounts of nephrotoxic serum (NTS), resulting in gradual increase in markers of kidney damage in serum and urine; n are number of mice per dose. (B) Transmission electron micrograph of kidney venules showing an example of intravascular occlusion in the NTS-treated mice, from 2 mice and 25–30 images analyzed. (C) Levels of the indicated metabolites in plasma of control and *Fgr*^{-/-} mice before and after induction of glomerulonephritis with LPS plus NTS. The control group was treated with LPS only; n are mice analyzed per group. (D–E) Mice reconstituted with marrow from wild-type or *Fgr*^{-/-} donor mice were infected with *C. albicans* (D) or *S. aureus* (E) and infection progression was measured by weight loss, and in the case of *C. albicans* infection by scoring the fungal load in kidneys (CFU); n are mice analyzed per group and data in (A) was analyzed by unpaired t-test. For (D–E) groups were compared by two-way ANOVA analyses for weight loss, and unpaired t-test for CFUs. Data in (C) was analyzed by one-way ANOVA with Tukey's multiple comparisons test. (F) Scheme modeling neutrophil states, transitions, and delivery of inflammatory signals to the host tissues from B3 cells. Each transition is proposed to be caused by different signals, e.g. delivered by platelets and PSGL-1 for initial transition from B1 to B2, and via Fgr for transitions from B2 to B3. The number of replicates (n) per group is indicated in each panel.

[Source data](#)

Supplementary information

Supplementary Information

This file contains Supplementary Tables 1–5 and captions for Supplementary Videos 1–6.

Reporting Summary

Supplementary Video 1

Multiphoton imaging of leukocytes in influenza-infected trachea.

Supplementary Video 2

Multiphoton imaging of leukocytes during ischaemia–reperfusion injury.

Supplementary Video 3

Multiphoton imaging of DCs and neutrophils during laser-burn injury.

Supplementary Video 4

Spinning-disk imaging of inflamed venules for behavioural analysis of intravascular neutrophils.

Supplementary Video 5

Three behavioural states for intravascular neutrophils within inflamed venules.

Supplementary Video 6

Filtering out subcellular fragments.

Source data

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

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

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

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

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

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

[**Source Data Extended Data Fig. 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

Crainiciuc, G., Palomino-Segura, M., Molina-Moreno, M. *et al.*
Behavioural immune landscapes of inflammation. *Nature* **601**, 415–421
(2022). <https://doi.org/10.1038/s41586-021-04263-y>

- Received: 10 June 2021
- Accepted: 18 November 2021
- Published: 05 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04263-y>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04263-y>.

- Article
- Open Access
- [Published: 05 January 2022](#)

RNA profiles reveal signatures of future health and disease in pregnancy

- [Morten Rasmussen](#) ORCID: [orcid.org/0000-0002-2743-5683¹](https://orcid.org/0000-0002-2743-5683),
- [Mitsu Reddy¹](#),
- [Rory Nolan¹](#),
- [Joan Camunas-Soler¹](#),
- [Arkady Khodursky¹](#),
- [Nikolai M. Scheller²](#),
- [David E. Cantonwine³](#),
- [Line Engelbrechtsen⁴](#),
- [Jia Dai Mi](#) ORCID: [orcid.org/0000-0001-5723-4252⁵](https://orcid.org/0000-0001-5723-4252),
- [Arup Dutta⁶](#),
- [Tiffany Brundage¹](#),
- [Farooq Siddiqui¹](#),
- [Mainou Thao¹](#),
- [Elaine P. S. Gee](#) ORCID: [orcid.org/0000-0001-7453-4433¹](https://orcid.org/0000-0001-7453-4433),
- [Johnny La¹](#),
- [Courtney Baruch-Gravett⁷](#),
- [Mark K. Santillan⁸](#),
- [Saikat Deb^{6,9}](#),
- [Shaali M. Ame⁹](#),
- [Said M. Ali⁹](#),
- [Melanie Adkins¹⁰](#),
- [Mark A. DePristo¹¹](#),
- [Manfred Lee¹](#),
- [Eugeni Namsaraev¹](#),
- [Dorte Jensen Gybel-Brask^{12,13}](#),
- [Lillian Skibsted¹²](#),
- [James A. Litch](#) ORCID: [orcid.org/0000-0003-0570-0308⁷](https://orcid.org/0000-0003-0570-0308),
- [Donna A. Santillan](#) ORCID: [orcid.org/0000-0002-6180-9714⁸](https://orcid.org/0000-0002-6180-9714),

- [Sunil Sazawal⁶](#),
- [Rachel M. Tribe](#) [ORCID: orcid.org/0000-0003-3675-9978⁵](#),
- [James M. Roberts](#) [ORCID: orcid.org/0000-0002-2671-3207¹⁴](#),
- [Maneesh Jain¹](#),
- [Estrid Høgdall¹³](#),
- [Claudia Holzman¹⁰](#),
- [Stephen R. Quake](#) [ORCID: orcid.org/0000-0002-1613-0809^{15,16,17}](#),
- [Michal A. Elovitz](#) [ORCID: orcid.org/0000-0001-7554-7180^{1,18}](#) &
- [Thomas F. McElrath](#) [ORCID: orcid.org/0000-0003-0857-0107³](#)

[*Nature*](#) volume 601, pages 422–427 (2022)

- 36k Accesses
- 1 Citations
- 1617 Altmetric
- [Metrics details](#)

Subjects

- [Gene expression](#)
- [Predictive markers](#)

Abstract

Maternal morbidity and mortality continue to rise, and pre-eclampsia is a major driver of this burden¹. Yet the ability to assess underlying pathophysiology before clinical presentation to enable identification of pregnancies at risk remains elusive. Here we demonstrate the ability of plasma cell-free RNA (cfRNA) to reveal patterns of normal pregnancy progression and determine the risk of developing pre-eclampsia months before clinical presentation. Our results centre on comprehensive transcriptome data from eight independent prospectively collected cohorts comprising 1,840 racially diverse pregnancies and retrospective analysis of 2,539 banked plasma samples. The pre-eclampsia data include 524 samples (72 cases and 452 non-cases) from two diverse independent cohorts collected 14.5 weeks (s.d., 4.5 weeks) before delivery. We show that cfRNA signatures from a single blood draw can track pregnancy progression at the placental, maternal and fetal levels and can robustly predict pre-eclampsia, with a sensitivity of 75% and a positive predictive value of 32.3% (s.d., 3%), which is superior to the state-of-the-art method². cfRNA signatures of normal pregnancy

progression and pre-eclampsia are independent of clinical factors, such as maternal age, body mass index and race, which cumulatively account for less than 1% of model variance. Further, the cfRNA signature for pre-eclampsia contains gene features linked to biological processes implicated in the underlying pathophysiology of pre-eclampsia.

[Download PDF](#)

Main

The period from conception to delivery represents the most rapid growth and development in an individual's life. The ability to support this development requires dramatic and poorly understood alterations in maternal physiology. Research into human pregnancy has clear ethical constraints, and the unique character of human gestation has limited deeper understanding of the physiology and pathophysiology of pregnancy³. Haemochorionic placentation is found among many mammalian species; however, in humans, it involves a unique degree of trophoblastic invasion^{4,5}, and because pre-eclampsia occurs predominantly in humans, conventional animal models are of limited value^{6,7}. Pre-eclampsia, a condition marked by maternal endothelial dysfunction and associated new-onset maternal hypertension, complicates up to 1 in 12 pregnancies and is a significant cause of maternal morbidity and higher lifetime risk of cardiovascular disease¹.

Here we demonstrate the ability of cfRNA transcripts to establish the normative responses of both maternal and fetal tissues characteristic of normal pregnancy progression. By implication, deviation from normative cfRNA expression patterns should allow the prediction of impending pathology before its presentation. We demonstrate the use of cfRNA to characterize women at risk of pre-eclampsia months before diagnosis. Notably, the cfRNA profiles identify risk solely through molecular mechanisms common to pre-eclampsia and are therefore exclusive of clinical variables such as race, body mass index (BMI), maternal comorbidities and/or obstetrical history.

In this study, we gather the largest and most diverse dataset of maternal transcriptomes to date. Samples were drawn from eight prospectively collected cohorts that provided $n = 2,539$ plasma samples from $n = 1,840$ pregnancies for women of multiple ethnicities, nationalities, geographic locations and socioeconomic contexts, while covering a range of gestational ages (Fig. 1a). The broad sociodemographic spectrum of our data (Table 1 and Supplementary Table 1) enabled us to test the applicability of maternal transcriptomes at one gestational time point. A detailed description of each cohort and the methodology is available in the [Supplementary Information](#).

Fig. 1: Overview of plasma sampling and cohorts and gestational age modelling.

 **figure 1**

a, Cohorts are labelled A–H (Table 1). Circles represent plasma samples from liquid biopsies ($n = 2,539$). Colours represent the race of the maternal donor. **b**, Model predictions from the hold-out test ($n = 474$) using cfRNA transcript data in the Lasso linear model versus ultrasound-predicted gestational age. The dark grey zone represents 1 s.d., and the light grey zone represents 2 s.d. **c**, Variance explained from ANOVA.

Table 1 Sample overview

RNA signal independent of clinical factors

Ultrasound-based gestational age has long been used as a surrogate measure of pregnancy progression. Here, we show that a cfRNA signature is as accurate a measure of gestational age while also providing insights into the biology of pregnancy progression. As a first step to develop a machine learning model, we divided our data from all full-term pregnancies without complications into a training set ($n = 1,908$ samples) and a test set ($n = 474$ samples), stratified by gestational age so that all age strata were represented proportionally. Before modelling, we standardized the means of gene counts across all cohorts ([Methods](#) and Extended Data Fig. 5). A Lasso linear model was fitted to predict gestational age in the training set, with a test set performance of a mean absolute error of 14.7 days (Fig. 1b, Extended Data Fig. 6 and Supplementary Data 1), referencing to first-trimester fetal ultrasound biometry. Overall, the error of our model is equivalent to that of second-trimester ultrasound and superior to that with third-trimester ultrasound⁸, and could provide an alternative dating procedure for women who start prenatal care later in pregnancy.

Next, we explored whether inclusion of clinical variables altered model performance. By analysis of variance (ANOVA), we showed that the model was driven almost entirely by information from the cfRNA transcripts, with BMI, maternal age and race accounting for less than 1% of variance (Fig. 1c). Rebuilding the gestational age model including maternal race, BMI and age provided no improvement in accuracy (0.07 days, not significant by bootstrap test).

Fetal signatures in maternal circulation

As the cfRNA signatures for gestational age demonstrated a dynamic change in transcripts as pregnancy progresses, we then explored whether transcripts found in the maternal circulation during pregnancy could be linked to their tissue of origin.

Specifically, we sought to ascertain whether the molecular status of the placenta, fetal organs and/or maternal tissues (cervix and/or uterus) could be assessed by examining cfRNA profiles. While fetal cells are known to pass into the maternal circulation^{9,10}, individual transcripts from the fetus or fetal cell types are relatively rare in maternal plasma; thus, we investigated these signals by analysing gene sets from Gene Ontology¹¹ or the Molecular Signatures Database^{12,13}. Using longitudinal data from cohort H covering 93 women sampled four times during pregnancy ([Supplementary Information](#)), we first confirmed that we could identify pregnancy-related sets such as those for gonadotropin and oestrogen pathways (Extended Data Fig. 1) and that the signal from the gestational age model increased with gestational age as did signal from the placenta (Fig. 2a, b and [Methods](#)). We show that hundreds of independently identified gene sets in maternal blood mirror the maternal and fetal physiological changes expected during pregnancy. Specifically, using single-cell RNA-seq data from adult and fetal organs (Supplementary Table 2), we were able to confirm changes in fetal gene sets, including those involved in fetal heart development, in maternal blood (Fig. 2c). Furthermore, the cfRNA profiles reflect expected changes in maternal tissues, such as the uterus and cervix, with progressively increasing expression of collagen and extracellular matrix gene sets¹⁴ (Fig. 2d). Extended Data Fig. 2 shows additional examples of fetal gene sets, including those of nephron progenitor cells for which expression become less abundant with gestational age in accordance with a decrease in the nephrogenic zone width^{15,16} and those in the gastrointestinal tract, where the oesophagus develops early with associated gene expression decreasing later versus small intestine where associated gene expression shows a steady increase¹⁷.

Fig. 2: Temporal profiles of pregnancy pathways for gene sets from the gestational age model and independently identified gene sets for placenta, developing fetal heart and collagen extracellular matrix known to be involved in uterus and cervix growth over gestation.

 **figure 2**

a–d, Maternal plasma transcriptome fractions for gene sets averaged across all samples in each collection window. Gestational age model (**a**), placenta (**b**), developing heart (**c**) and collagen extracellular matrix (ECM) (**d**). Error bars correspond to the 95% confidence interval around the mean. CPM, counts per million. $n = 93$ for each time point and gene set. **e–h**, Signal across all cohorts with longitudinal data: gestational age model (**e**), placenta (**f**), developing heart (**g**) and collagen ECM (**h**). Linear fits are shown of transcriptome fractions for all samples across corresponding gestational ages recorded at collection times. The band around the solid line corresponds to the 95% confidence interval. All slopes for the gestational age coefficients are distinct from 0 at a confidence level of 0.05. Cohort is indicated by colour.

To test whether the identified gene sets were uniquely associated with pregnancy progression, we next compared the observed gestational age collection time labels to a set of randomly permuted collection time labels. This comparison verified that all selected gene sets were associated with pregnancy progression (Extended Data Fig. 3). The directional signals could be confirmed in three independent cohorts ($n = 351$ women) for which longitudinal data were available (Fig. 2e–h). In all cases, the slopes for the gestational age coefficients were distinct from 0 at a 0.05 confidence level. In total, we tested 793 gene sets from single-cell analyses^{12,13}, comprising 384 gene sets from adult and 409 gene sets from fetal tissues. Of these, 129 gene sets (55 fetal) were significantly correlated with gestational age, of which 99 gene sets (40 fetal) showed increased signal and 30 gene sets (15 fetal) showed decreased signal as a function of gestational age at collection in cohort H, and were confirmed in at least two other cohorts with longitudinally sampled individuals (Supplementary Data 2). As changes in these predefined gene sets were only significant in the context of gestational age across at least three cohorts with longitudinal information, we present here a non-invasive window into maternal–fetal development from a maternal blood sample.

Early prediction of pre-eclampsia

Having established that cfRNA profiles can reveal and characterize molecular changes in the maternal–placental–fetal unit over gestation, it is likely that disruption of these pathways might identify women at risk for adverse pregnancy outcomes such as pre-eclampsia.

We evaluated the ability of cfRNA signatures in maternal blood, during the second trimester (16–27 weeks), to predict the development of pre-eclampsia. Maternal blood draws occurred, on average, 14.5 weeks (s.d., 4.5 weeks) before delivery (Fig. [3a](#)); in contrast to work by Munchel et al.^{[18](#)} where plasma was collected at the time of diagnosis, the gestational age time points in our analysis correspond to timepoints where women are asymptomatic. A case–control study with 72 cases of pre-eclampsia and 452 non-cases selected from two independent cohorts (cohorts A and E) was performed ([Supplementary Information](#)). Cohort E included 31 controls with chronic hypertension and 19 controls with gestational hypertension and both cohorts included spontaneous preterm birth samples along with the normotensive term controls. Pre-eclampsia was defined by criteria consistent with those from the 2013 Task Force on Hypertension in Pregnancy (ACOG 2013), and each case was adjudicated by two board-certified physicians. As before, a cohort correction was applied before modelling.

Fig. 3: Features and model performance for prediction of pre-eclampsia.

 **figure 3**

a, Sample collection time (dashed lines) and delivery time (solid lines) for women with pre-eclampsia (purple and green) and controls (grey). Gradients illustrate timelines for developing pathophysiology and onset of clinical symptoms. **b**, Quantile–quantile plot of ranked Spearman P values for women with pre-eclampsia (cases) versus controls. P values were calculated from Spearman correlation on cohort-corrected data for each gene. The genes used in the model are labelled. The black dotted line represents the expectation. **c**, Receiver operating characteristic curve (mean and 95% confidence interval) for the logistic regression model for pre-eclampsia ($n = 524$). **d**, Kaplan–Meier curves of deliveries in test-positive and test-negative populations ($n = 439$), excluding spontaneous preterm deliveries.

Two-sided Spearman correlation tests identified signatures that separated the cases and controls; in each round of cross-validation, we retained features with an adjusted P value below 0.05 ([Methods](#)) and consistently identified seven genes: *CLDN7*, *PAPPA2*, *SNORD14A*, *PLEKHH1*, *MAGEA10*, *TLE6* and *FABP1* (Fig. [3b](#)).

Four of the genes selected for modelling have functions relevant to pre-eclampsia or placental development. *PAPPA2*, encoding pregnancy-associated plasma protein 2, is

expressed in the placenta¹⁹, specifically in trophoblast cells. It has previously been linked to the development of pre-eclampsia and has been associated with inhibition of trophoblast migration, invasion and tube formation^{20,21}. Claudin 7 (*CLDN7*) is involved in tight cell junction formation and blastocyst implantation; in healthy pregnancies, expression of *CLDN7* is reduced in response to oestrogen at the time of implantation^{22,23}. Similarly, *TLE6* has also been linked to preimplantation and early embryonic lethality²⁴. Fatty acid-binding protein 1 (*FABP1*) was first purified from human cytotrophoblasts and is known to be highly expressed in the fetal liver; it is critical for fatty acid uptake and transport²⁵ and is upregulated threefold when cytotrophoblasts differentiate to syncytiotrophoblasts at implantation²⁶. The other three genes that make up the pre-eclampsia cfRNA signature (*SNORD14A*, *PLEKHH1* and *MAGEA10*) have been associated with pre-eclampsia through bioinformatic analyses, although their function is less well understood^{27,28}. Two of the identified genes, *PAPPA2* and *FABP1*, were also identified in the gestational age model and highlight the imbalance in cfRNA signatures between pregnancy progression and pathology.

On the basis of these identified gene features, a logistic regression model in a leave-one-out cross-validation set-up was used to estimate the probability of pre-eclampsia. This model framework was chosen on the basis of learning curve analyses ([Methods](#) and Extended Data Fig. [7](#)). At a sensitivity of 75%, our cfRNA model achieved a positive predictive value (PPV) of 32.3% (s.d., 3%) given a prevalence of pre-eclampsia of 13.7% in our study, superior to PPVs reported from current clinical state-of-the-art models, which are driven largely by maternal factors²; the area under the curve (AUC) for the model was 0.82 (95% confidence interval, ± 0.06 ; Fig. [3c](#)). Consistent with our findings with the gestational age model, inclusion of clinical variables (maternal BMI, age and race) had no effect on performance, as the classifier assigns zero weight to these clinical variables and they explain <1% of the variance based on ANOVA analyses. The lack of contribution to cfRNA profiles from clinical factors highlights the generalizability of these profiles to diverse populations.

When comparing gestational age at delivery between test-positive and test-negative individuals, a significant shift was found in the timing of delivery, with the test-positive population delivering earlier during gestation ($P < 2 \times 10^{-7}$; Fig. [3d](#)). A positive test correctly identified 73% of individuals destined to have a medically indicated preterm birth over 3 months in advance of the onset of clinical symptoms or delivery.

To further understand molecular signature changes and how they might reflect the pathophysiology driving pre-eclampsia, we performed pathway analysis. The top upregulated pathways were dominated by structural cell functions, including placental blood vessel development, artery morphogenesis and embryonic placental

development (Extended Data Fig. 4a), while the majority of downregulated pathways were related to immune pathways (Extended Data Fig. 4b). Both the upregulated and downregulated gene sets aligned with the accepted mechanism of pathogenesis for pre-eclampsia²⁹.

In cohort E, the non-case group contained both normotensive women ($n = 263$) and women with chronic ($n = 31$) or gestational ($n = 19$) hypertension. Genes identified through comparison of the groups with chronic or gestational hypertension with the normotensive group showed no overlap with genes significant for pre-eclampsia (two-sided Spearman correlation test, $P < 0.05$). Additionally, no genes were differentially expressed in the chronic or gestational hypertensive groups when compared with the normotensive group. While others have published studies designed to determine the effect of hypertension more generally on gene expression (e.g., Zeller et al.³⁰), here, we demonstrate that the signal for pre-eclampsia is specific to hypertension driven by a placental disorder and the signature is independent of signals associated with chronic hypertension. Clinically, it can be quite challenging to differentiate superimposed pre-eclampsia in women with pre-existing hypertension from exacerbation of baseline chronic hypertension. This difference is important, as one requires delivery for cure while the other usually does not.

As pre-eclampsia and spontaneous preterm birth are theorized to have some overlapping molecular pathways^{31,32}, we tested whether excluding non-case samples with deliveries before gestational week 37 ($n = 85$) would affect test prediction. Removal of spontaneous preterm delivery samples did not alter the performance of the model ($AUC = 0.79$; 95% confidence interval, ± 0.06), suggesting that inclusion of spontaneous preterm birth samples in the non-case group does not affect the pre-eclampsia classifier.

We report a standalone molecular predictor that has the potential to be an early detector of pre-eclampsia with a PPV of 32% that is based entirely on transcripts and is exclusive of clinical variables. This predictor contrasts with state-of-the-art methods, which are dependent on clinical factors and achieve a PPV of 4.4%².

Discussion

While other studies have looked at circulating biomarkers, a recent comprehensive review³³ concluded that more data early in pregnancy are needed to support clinical value. Here, we reveal the ability of cfRNA transcripts to provide comprehensive molecular profiles of pregnancy progression by including signals from the placenta and the fetus. We have shown that novel transcript signatures from a single blood sample can (1) accurately track pregnancy progression independently of clinical factors and (2) reliably identify women at risk of developing pre-eclampsia months

before presentation of the disease. Given the large sample size and diversity in our study population, it is noteworthy that race has a negligible effect on the expression patterns of gestational age estimates and pre-eclampsia risk evaluation. These findings allow for the development of personalized assessments for pregnancy.

Equally important, our work allows for the assessment of maternal risk independently of clinical factors, such as race, that are fraught with bias. The inclusion of race in clinical assessments results in miscalculation of patient risk and underdiagnoses^{34,35,36}. While we acknowledge that, within specific subpopulations, the prevalence of complications such as pre-eclampsia may be higher, the evaluation of cfRNA transcripts directly exposes the developing pathophysiology. Further research will be needed to identify drivers of the identified pathophysiological pathways; the focus on molecular mechanisms allows stratification of risk without the need for enrichment of ‘pretest’ probabilities based on maternal sociodemographic characteristics. Further, an understanding of the maternal–fetal–placental transcriptome also represents a vehicle by which comprehension of the biological underpinnings of maternal–fetal development can be improved and provides novel insights into interactions across the maternal–fetal dyad. This holds the promise of precision therapeutic interventions that can target molecular subtypes of pre-eclampsia and preterm birth.

Improvement in maternal outcomes has been limited by the inability to access pregnancy tissues and a lack of understanding of the specific molecular phenotypes that identify those at risk before onset of symptoms. Our findings can now be leveraged to more accurately provide information on future maternal and fetal health and disease. Thus, our approach opens new therapeutic windows to effectively decrease maternal and neonatal morbidity and mortality.

Methods

The Mirvie RNA technology

cfRNA isolation

Plasma samples received on dry ice from our collaborators were stored at –80 °C until further processing. Total circulating nucleic acid was extracted from plasma ranging in volume from ~215 µl to 1 ml, using a column-based commercially available extraction kit, following the manufacturer’s instructions (Plasma/Serum Circulating and Exosomal RNA purification kit, Norgen, 42800).

Following extraction, cfDNA was digested using Baseline-ZERO DNase (Epicentre) and the remaining cfRNA was purified using an RNA Clean and Concentrator-5 kit

(Zymo, R1016) or an RNeasy MinElute Cleanup kit (Qiagen, 74204).

RT–qPCR assay

We performed PCR with reverse transcription (RT–qPCR) analysis to assess the relative amount of cfRNA extracted from each sample. We measured and compared the threshold cycle (C_t) values from each RNA sample using a three-colour multiplex qPCR assay from the TaqPath 1-Step Multiplex Master Mix kit (ThermoFisher Scientific, A28526) and a Quant Studio 5 system. We also measured the C_t values for an endogenous housekeeping gene (*ACTB*; ThermoFisher Scientific, 4351368).

cfRNA library preparation

cfRNA libraries were prepared using the SMARTer Stranded Total RNAseq-Pico Input Mammalian kit (Takara, 634418) following the manufacturer’s instructions, except that we did not use ribo depletion. Library quality was assessed by RT–qPCR following the method described for assessing RNA measurements and fragment analysis on a Fragment Analyzer 5300 (Agilent Technologies).

Enrichment and sequencing

Libraries were normalized before pooling for target capture. We used a SureSelect Target Enrichment kit (Agilent Technologies, 5190-8645) and followed the manufacturer’s instructions for hybrid capture. Samples were quantified, and 50-bp, paired-end sequencing was performed on a Novaseq S2. Between 96 and 144 samples were pooled and sequenced per sequencing run.

Analysis for outliers

qPCR of *ACTB* as well as MultiQC sequencing metrics were monitored to eliminate sample outliers before performing gene expression analyses. Individual samples more than 3 s.d. from the mean were removed as outliers. A total of 193 of 2,732 samples (7.1%) were removed following this filtering.

Read processing

Reads were processed following a similar protocol to that reported in Ngo et al.³⁷. Briefly, raw sequencing reads were trimmed using trimmomatic³⁸ and then mapped to hg38 using the STAR aligner³⁹. After removing duplicates using Picard tools, gene counts were generated with htseq⁴⁰.

Cohort correction and feature normalization

For each gene, its relationship to total counts per sample was measured and corrected using linear model residuals. Extended Data Fig. [5a](#), [b](#) shows what this looks like for the gene *ACTB*.

We also sought to correct the genes such that each cohort had the same mean value for each gene. However, the cohorts came from different parts of the gestational age spectrum. Therefore, only cohort effects orthogonal to the gestational age effect were corrected. This is shown in Extended Data Fig. [5c](#), [d](#) for the gene *CAPN6*. Each cohort was given its own colour.

Cohort E (bright yellow) had unusually low counts for its gestational age range before correction, and this effect was removed by correction.

Using principal-component analysis (PCA) to compress the high-dimensional space of all genes, the correction could be seen to clarify the separation of samples by gestational age as indicated by the colour gradient (Extended Data Fig. [5e](#), [f](#)).

Linear correction algorithm

1. In the training, correct for (remove the effect of) the variable(s) of interest (e.g., gestational age) using linear model residuals.
2. Learn the required correction for the variables you wish to correct for in this corrected training data.
3. The residuals of that model (in the raw training and testing data) are your corrected data.

Note: the correction was learned entirely in the training data and the variable of interest in the testing data was never used, negating the possibility of a data leak.

Lasso linear model for gestational age prediction and ANOVA

The Lasso model used in the gestational age model had its parameters chosen via 10-fold cross-validation in the training set. The largest cross-validation score within one standard error of the best cross-validation score was chosen (Breiman strategy). We limited our feature space by excluding pseudogenes and non-coding genes, as well as genes with median expression greater than zero, leaving a total of 13,208 features to evaluate. A final Lasso with this was then trained on the whole training set and evaluated in the test set. This was all done with the `glmnet` R package using the `cv.glmnet()` function.

The model uses 674 of the available gene features (Supplementary Data 1), although this includes a long tail of features with low contribution. We tested performance for the 50 most informative features from the model and obtained a mean absolute error of 15.4 days. The continued reduction in error as we reached our complete training set of $n = 1,908$ samples indicated that model learning was not exhausted and that additional samples would have increased performance (Extended Data Fig. 6). Notably, as seen in Extended Data Fig. 6, the similar performance in cross-validation and on the independent held-out test data indicated that the model was not overfit with the 674 gene features. To determine how far the model could be extrapolated, a final model was built using all data; this gave a mean absolute error of 13 days across the entire dataset.

Gestational age learning curve

The main gestational age modelling was done with an 80/20 train/test split. To assess model performance after decreasing amounts of training data, one can repeat analyses with 70/30 splits, 60/40 splits and so on (doing so repeatedly with different random splits to quantify uncertainty). In this way, one builds a learning curve (Extended Data Fig. 6) with different training set sizes on the x axis and model performance on the y axis.

Gestational age model without cohort correction

For this approach, we selected all samples from healthy pregnancies and split the dataset into a training set (80% of data) and a test set (20% of data), in which samples were stratified by cohort. Samples that did not pass quality-control filtering based on basic sequencing metrics had been previously excluded from analysis. We trained a Lasso model to predict the gestational age at collection for each sample using the mean absolute error as an optimization metric and 10-fold cross-validation in the training set. We used all genes with mean $\log_2(\text{counts per million (CPM)} + 1) > 1$ (12,921 genes) plus a set of sequencing metrics as features for training. Modelling was performed in $\log_2(\text{CPM} + 1)$ space, and all data were centred and scaled before modelling using the training set statistics. This led to a model with a mean absolute error of 15.9 days in the withheld test set using 487 transcriptomic features. We then selected the top 53 features of this model and retrained the Lasso using the same approach described above, achieving a mean absolute error of 16.6 days in the held-out test set.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA)^{11,41} was done with the fast GSEA algorithm⁴² using Bioconductor's fgsea package⁴³. Gene sets were compiled from the Molecular Signatures Database (MSigDB)^{11,12} using the CRAN msigdbr v7.2 API and directly from c8.all.v7.3.symbols.gmt. We focused on two collections of gene sets: the Gene Ontology (GO) subcollection of the ontology gene sets, C5:GO, and the cell type signature gene sets, C8 v7.3. Genes were ranked on the basis of their shrunken log-transformed fold change values and associated Wald test *P* values obtained from analysis of differential expression using Bioconductor's DESeq2 (ref. ⁴⁴), represented as $-\log_{10}(P \text{ value}) \times \text{shrunkenLFC}$. GSEA was carried out on 372 samples from cohort H collected from 93 women with healthy pregnancies over four draw intervals during pregnancy, 11.4–14 weeks, 18–21 weeks, 22.8–27.8 weeks and 29.2–34.8 weeks. Shrunken log-transformed fold change values and corresponding *P* values were obtained from all six pairwise contrasts between the four draws. We used 102 fetal gene sets that were significantly enriched (Benjamini–Hochberg adjusted *P*<0.01) in at least one pairwise comparison (Supplementary Table ²) in downstream analyses, including analysis of plasma transcriptome partitioning and set-specific longitudinal trends.

Using a GO collection of gene sets, we validated our approach and identified seven pregnancy-related sets that were significantly enriched in the comparison between early- and late-pregnancy samples (Extended Data Figure ¹). Three gene sets in the gonadotropin and oestrogen pathways exhibited significant changes consistent with known physiology⁴⁵.

Evaluating changes in plasma transcriptome partitioning

The plasma transcriptome can be phenomenologically viewed as being partitioned into characteristic sets of genes. We assessed this partitioning in each cfRNA sample by converting raw gene counts to CPM and summing CPM over all genes in each of the sets. The resulting cumulative CPM score, which is a relative measure of the abundance of each gene set in the overall transcriptome, was used to directly compare gene sets across collection time points. Cumulative CPM scores for all gene sets significantly enriched between collections 1 and 4 were calculated for every cfRNA sample. The scores for each sample were regressed onto the recorded gestational age (in weeks) using a linear model. Gene sets with an adjusted *P* value for the gestational age coefficient <0.01 were considered as having a significant (positive or negative) trend in their relative abundance. The association of these trends with the time component in the data was further verified by scrambling the temporal structure and re-examining the trends along the original time variable. For each mother, we also evaluated the monotonicity of the cumulative CPM score function along the collection times. Because there are 24 possible permutations of order for the four collection times and only one of those permutations allows for a monotonic upward trend (with one for

a downward trend), we were able to analytically assess the significance of the observed number of monotonic trends among 93 mothers using a chi-squared test.

Pre-eclampsia analysis and learning curve

CIs for AUCs and sensitivity, specificity and PPV were all found via bootstrapping. PPV was calculated as $PPV = (\text{sensitivity} \times \text{prevalence}) / ((\text{sensitivity} \times \text{prevalence}) + ((1 - \text{specificity}) \times (1 - \text{prevalence})))$.

To build the learning curve (Extended Data Fig. 7), we increased the size of the training set going from two- to ninefold cross-validation with a constant model: logistic regression with gene features chosen by Spearman correlation tests with an adjusted P -value threshold of 0.05. The point on the right connected to the learning curve via a dashed line is the leave-one-out cross-validation result shown in the main text.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Data are available with a signed data use agreement to protect identifiable data; please contact research@mirvie.com.

Code availability

Code is available as three packages in the following repositories: mirmisc, <https://doi.org/10.5281/zenodo.5604683>; mirmodels, <https://doi.org/10.5281/zenodo.5593282>; and mirr, <https://doi.org/10.5281/zenodo.5593280>.

References

1. 1.

Rich-Edwards, J. W., Fraser, A., Lawlor, D. A. & Catov, J. M. Pregnancy characteristics and women's future cardiovascular health: an underused opportunity to improve women's health? *Epidemiol. Rev.* **36**, 57–70 (2014).

2. 2.

Tan, M. Y. et al. Screening for pre-eclampsia by maternal factors and biomarkers at 11–13 weeks' gestation: first-trimester PE screening. *Ultrasound Obstet. Gynecol.* **52**, 186–195 (2018).

3. 3.

Marinić, M. & Lynch, V. J. Relaxed constraint and functional divergence of the progesterone receptor (PGR) in the human stem-lineage. *PLoS Genet.* **16**, e1008666 (2020).

4. 4.

Robillard, P.-Y., Dekker, G. A. & Hulsey, T. C. Evolutionary adaptations to pre-eclampsia/eclampsia in humans: low fecundability rate, loss of oestrus, prohibitions of incest and systematic polyandry. *Am. J. Reprod. Immunol.* **47**, 104–111 (2002).

5. 5.

McCarthy, F. P., Kingdom, J. C., Kenny, L. C. & Walsh, S. K. Animal models of preeclampsia; uses and limitations. *Placenta* **32**, 413–419 (2011).

6. 6.

Chez, R. A. Nonhuman primate models of toxemia of pregnancy. *Perspect. Nephrol. Hypertens.* **5**, 421–424 (1976).

7. 7.

Malassiné, A., Frendo, J. L. & Evain-Brion, D. A comparison of placental development and endocrine functions between the human and mouse model. *Hum. Reprod. Update* **9**, 531–539 (2003).

8. 8.

Skupski, D. W. et al. Estimating gestational age from ultrasound fetal biometrics. *Obstet Gynecol* **130**, 433–441 (2017).

9. 9.

Khosrotehrani, K., Johnson, K. L., Cha, D. H., Salomon, R. N. & Bianchi, D. W. Transfer of fetal cells with multilineage potential to maternal tissue. *JAMA* **292**,

75–80 (2004).

10. 10.

Kahn, D. A. & Baltimore, D. Pregnancy induces a fetal antigen-specific maternal T regulatory cell response that contributes to tolerance. *Proc. Natl Acad. Sci. USA* **107**, 9299–9304 (2010).

11. 11.

Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

12. 12.

Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

13. 13.

Liberzon, A. et al. Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

14. 14.

Shi, J.-W. et al. Collagen at the maternal-fetal interface in human pregnancy. *Int. J. Biol. Sci.* **16**, 2220–2234 (2020).

15. 15.

Menon, R. et al. Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development* **145**, dev164038 (2018).

16. 16.

Ryan, D. et al. Development of the human fetal kidney from mid to late gestation in male and female infants. *EBioMedicine* **27**, 275–283 (2018).

17. 17.

Gao, S. et al. Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat. Cell Biol.* **20**, 721–734 (2018).

18. 18.

Munchel, S. et al. Circulating transcripts in maternal blood reflect a molecular signature of early-onset preeclampsia. *Sci. Transl. Med.* **12**, eaaz0131 (2020).

19. 19.

Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

20. 20.

Kramer, A. W., Lamale-Smith, L. M. & Winn, V. D. Differential expression of human placental PAPP-A2 over gestation and in preeclampsia. *Placenta* **37**, 19–25 (2016).

21. 21.

Chen, X et al. The potential role of pregnancy-associated plasma protein-A2 in angiogenesis and development of preeclampsia. *Hypertens. Res.* **42**, 970–980 (2019).

22. 22.

Poon, C. E., Madawala, R. J., Day, M. L. & Murphy, C. R. Claudin 7 is reduced in uterine epithelial cells during early pregnancy in the rat. *Histochem. Cell Biol.* **139**, 583–593 (2013).

23. 23.

Schumann, S., Buck, V. U., Classen-Linke, I., Wennemuth, G. & Grüninger, R. Claudin-3, claudin-7, and claudin-10 show different distribution patterns during decidualization and trophoblast invasion in mouse and human. *Histochem. Cell Biol.* **144**, 571–585 (2015).

24. 24.

Alazami, A. M. et al. *TLE6* mutation causes the earliest known human embryonic lethality. *Genome Biol.* **16**, 240 (2015).

25. 25.

Wang, G., Bonkovsky, H. L., de Lemos, A. & Burczynski, F. J. Recent insights into the biological functions of liver fatty acid binding protein 1. *J. Lipid Res.* **56**,

2238–2247 (2020).

26. 26.

Cunningham, P. & McDermott, L. Long chain PUFA transport in human term placenta. *J. Nutr.* **139**, 636–639 (2009).

27. 27.

Ren, Z. et al. Distinct molecular processes in placentae involved in two major subtypes of preeclampsia. Preprint at *bioRxiv* <https://doi.org/10.1101/787796> (2019).

28. 28.

Gormley, M. et al. Preeclampsia: novel insights from global RNA profiling of trophoblast subpopulations. *Am. J. Obstet. Gynecol.* **217**, 200.e1–200.e17 (2017).

29. 29.

Redman, C. W. & Sargent, I. L. Latest advances in understanding preeclampsia. *Science* **308**, 1592–1594 (2005).

30. 30.

Zeller, T. et al. Transcriptome-wide analysis identifies novel associations with blood pressure. *Hypertension* **70**, 743–750 (2017).

31. 31.

Challis, J. R. et al. Inflammation and pregnancy. *Reprod. Sci.* **16**, 206–215 (2009).

32. 32.

Raghupathy, R. & Kalinka, J. Cytokine imbalance in pregnancy complications and its modulation. *Front. Biosci.* **13**, 985–994 (2008).

33. 33.

Carbone, I. F. et al. Circulating nucleic acids in maternal plasma and serum in pregnancy complications: are they really useful in clinical practice? A systematic review. *Mol. Diagn. Ther.* **24**, 409–431 (2020).

34. 34.

Vyas, D. A., Eisenstein, L. G. & Jones, D. S. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *N. Engl. J. Med.* **383**, 874–882 (2020).

35. 35.

Delgado, C. et al. Reassessing the inclusion of race in diagnosing kidney diseases: an interim report from the NKF-ASN Task Force. *J. Am. Soc. Nephrol.* **32**, 1305–1317 (2021).

36. 36.

Grobman, W. A. et al. Prediction of vaginal birth after cesarean delivery in term gestations: a calculator without race and ethnicity. *Am. J. Obstet. Gynecol.* <https://doi.org/10.1016/j.ajog.2021.05.021> (2021).

37. 37.

Ngo, T. T. M. et al. Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science* **360**, 1133–1136 (2018).

38. 38.

Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

39. 39.

Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

40. 40.

Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

41. 41.

Mootha, V. K. et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

42. 42.

Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at *bioRxiv* <https://doi.org/10.1101/060012> (2016).

43. 43.

Cre, A. S. Fast gene set enrichment analysis.
<https://doi.org/10.18129/B9.BIOC.FGSEA> (Bioconductor, 2017).

44. 44.

Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

45. 45.

Tal, R. & Taylor, H. S. Endocrinology of pregnancy. *Endotext* www.endotext.org (MDText.com, 2021).

Acknowledgements

We thank all women who donated blood samples and made this study possible. This research was conducted using specimens and data collected, stored and managed by INSIGHT, LIFECODES, The Women's Health Tissue, Pregnancy Outcomes and Community Health (POUCH), Prenatal Exposures and Preeclampsia Prevention (PEPP), Global Alliance to Prevent Prematurity and Stillbirth (GAPPS), Pemba Pregnancy and Newborn Discovery Cohort (PPNDC) and Roskilde biorepositories. We thank the Precia Group for introducing and coordinating with key study collaborators. Samples from the INSIGHT study were collected with support from Tommy's Charity (no. 1060508), the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) based at Guy's and St Thomas' National Health Service Foundation Trust, the Rosetrees Trust (charity no. 298582) (M303-CD1) and an NIHR Doctoral Research Fellowship (DRF-2013-06-171) to N.L. Hezelgrave. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. Research reported in this publication was supported by UI BioShare, the enterprise biospecimen management system supported by the University of Iowa's Carver College of Medicine, Holden Comprehensive Cancer Center and Institute for Clinical and Translational Science.

Author information

Affiliations

1. Mirvie, Inc., South San Francisco, CA, USA

Morten Rasmussen, Mitsu Reddy, Rory Nolan, Joan Camunas-Soler, Arkady Khodursky, Tiffany Brundage, Farooq Siddiqui, Mainou Thao, Elaine P. S. Gee, Johnny La, Manfred Lee, Eugeni Namsaraev, Maneesh Jain & Michal A. Elovitz

2. Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark

Nikolai M. Scheller

3. Brigham and Women's Hospital, Boston, MA, USA

David E. Cantonwine & Thomas F. McElrath

4. Department of Obstetrics and Gynecology, Herlev University Hospital, Herlev, Denmark

Line Engelbrechtsen

5. Department of Women and Children's Health, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London, St Thomas' Hospital Campus, London, UK

Jia Dai Mi & Rachel M. Tribe

6. Center for Public Health Kinetics, New Delhi, India

Arup Dutta, Saikat Deb & Sunil Sazawal

7. Global Alliance to Prevent Prematurity and Stillbirth (GAPPS), Lynnwood, WA, USA

Courtney Baruch-Gravett & James A. Litch

8. Department of Obstetrics & Gynecology, University of Iowa Hospitals & Clinics, Iowa City, IA, USA

Mark K. Santillan & Donna A. Santillan

9. Public Health Laboratory-Idc, Pemba, Zanzibar, Tanzania

Saikat Deb, Shaali M. Ame & Said M. Ali

10. Michigan State University, East Lansing, MI, USA

Melanie Adkins & Claudia Holzman

11. BigHat Biosciences, Inc., San Mateo, CA, USA

Mark A. DePristo

12. Department of Obstetrics, Zealand University Hospital, Roskilde, Denmark

Dorte Jensen Gybel-Brask & Lillian Skibsted

13. Department of Pathology, Herlev University Hospital, Herlev, Denmark

Dorte Jensen Gybel-Brask & Estrid Høgdall

14. Magee-Womens Research Institute, Department of Obstetrics and Gynecology and Reproductive Sciences, Epidemiology and Clinical and Translational Research University of Pittsburgh, Pittsburgh, PA, USA

James M. Roberts

15. Department of Bioengineering, Stanford University, Stanford, CA, USA

Stephen R. Quake

16. Chan Zuckerberg Biohub, Stanford, CA, USA

Stephen R. Quake

17. Department of Applied Physics, Stanford University, Stanford, CA, USA

Stephen R. Quake

18. Maternal and Child Health Research Program, Department of Obstetrics and Gynecology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Michal A. Elovitz

Contributions

M. Rasmussen, M.L., E.N., M.J., S.R.Q. and T.F.M. conceptualized and designed the study with input from the remaining authors. N.M.S., D.E.C., L.E., J.D.M., A.D., C.B.-G., M.K.S., S.D., S.M. Ame, S.M. Ali, M.A., D.J.G.-B., L.S., J.A.L., D.A.S., S.S., R.M.T., J.M.R., E.H., C.H. and T.F.M. provided samples and data to the study, curated the collection and obtained approvals for use in this study where required. M. Rasmussen, M. Reddy, J.C.-S. and E.N. designed laboratory protocols; all laboratory experiments were carried out by M. Reddy, T.B., M.T. and J.L. M. Rasmussen, R.N., J.C.-S., A.K., F.S., E.P.S.G., M.D., E.N. and S.R.Q. conceptualized computational analyses; R.N., J.C.-S., A.K., F.S. and E.P.S.G. implemented and reviewed code. M. Rasmussen, M.J., M.A.E. and T.F.M. drafted the manuscript with critical input from all authors.

Corresponding authors

Correspondence to [Morten Rasmussen](#), [Michal A. Elovitz](#) or [Thomas F. McElrath](#).

Ethics declarations

Competing interests

M. Rasmussen, M. Reddy, R.N., J.C.-S., A.K., T.B., F.S., M.T., E.P.S.G., J.L., M.L., E.N., M.J., M.A.E., M.D., S.R.Q. and T.M. have an equity interest in Mirvie. All cohort contributors were compensated for sample collection and/or shipping. T.M. serves on the scientific advisory board for Mirvie, NxPrenatal, Momenta Pharmaceuticals and Hoffmann–La Roche. M. Rasmussen, M. Reddy, R.N., J.C.-S., A.K., T.B., F.S., M.T., E.P.S.G., J.L., M.L., E.N., M.J., M.A.E., S.R.Q., M.K.S. and D.A.S. are inventors on patent applications (US20170145509A1, US9937182B2 and EP2954324A1) that cover the detection, diagnosis or treatment of pregnancy complications.

Additional information

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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

Extended data figures and tables

Extended Data Fig. 1 Temporal profiles of pregnancy-related endocrine signatures during pregnancy.

Seven pregnancy-related gene ontology term signatures identified as highly significantly enriched ($\alpha=0.01$) were profiled across collection times using cumulative CPM. Plasma transcriptome fractions for each gene set were averaged across all samples in each collection window with error bars corresponding to the 95% confidence interval around the mean. Panels correspond to different ranges of CPM, for the ease of comparison. CPM, counts per million. N=93 for each timepoint and gene set.

Extended Data Fig. 2 Temporal profiles of fetal gene sets from developing kidney and gastrointestinal tract.

a-c, Maternal plasma transcriptome fractions for gene sets averaged across all samples in a given collection window. Error bars correspond to the 95% confidence interval around the mean. CPM, counts per million. N=93 for each timepoint and gene set. **d-f**, signal across all cohorts with longitudinal data. Linear fits of transcriptome fractions for all samples across corresponding gestational ages recorded at the collection times. The band around the solid line corresponds to the 95% CI. All slopes for the gestational age coefficient are distinct from 0 at a confidence level of 0.05. Cohort is indicated by color.

Extended Data Fig. 3 Bootstrapping with and without time-scrambling.

a-d, for each of the significantly enriched gene sets, the trends were evaluated by bootstrapping (B=1,000) the original data (blue lines) and time-scrambled data (grey lines) obtained by reshuffling collection times.

Extended Data Fig. 4 Gene set enrichment analysis of preeclampsia for gene ontology (GO) gene sets.

a, Top-20 significantly upregulated gene sets. **b**, Top-20 significantly downregulated gene sets. Color gradient for adjusted p -value. NES, absolute normalized enrichment score.

Extended Data Figure 5 Effect of correcting for total count and cohort.

Counts for ACTB as a function of total counts for the sample before (**a**) and after (**b**) correction. Counts for CAPN6 as a function of gestational age for all samples used in

the gestational age model before (**c**) and after (**d**) cohort correction. Plot of first two principal components before (**e**) and after (**f**) cohort correction.

Extended Data Fig. 6 Learning curve for gestational age model.

Model for gestational age is trained with increasing sample size, error is plotted for both training set (Cross-validated, purple) and held-out test set (green). Error bars are 1 standard deviation.

Extended Data Fig. 7 Learning curve for preeclampsia model.

Model performance as a function of training set size. Error bars are 1 standard deviation.

Supplementary information

Supplementary Information

This file contains a cohort overview, supplementary analyses, additional references and Supplementary Tables 1 and 2.

Reporting Summary

Peer Review Information

Supplementary Data 1

Full list of features selected by the Lasso for the gestational age model, including their weights in the model.

Supplementary Data 2

Full list of gene sets with significant negative correlation to gestational age in at least three independent cohorts. Gene sets were discovered in cohort H and confirmed in at least two other cohorts (A, B or G). The slope is the average change in gene set expression in CPM per week. Adjusted *P* value for discovery cohort. Number of cohorts with significant signal, including discovery cohort. Type, origin of dataset, fetal or adult tissue. Additional sheets for component genes for each set of adult or foetal genes.

Supplementary Data 3

Full list of gene sets with significant positive correlation to gestational age in at least three independent cohorts. Gene sets were discovered in cohort H and confirmed in at least two other cohorts (A, B or G). The slope is the average change in gene set expression in CPM per week. Adjusted *P* value for discovery cohort. Number of cohorts with significant signal, including discovery cohort. Type, origin of dataset, fetal or adult tissue. Additional sheets for component genes for each set of adult or fetal genes.

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

Rasmussen, M., Reddy, M., Nolan, R. *et al.* RNA profiles reveal signatures of future health and disease in pregnancy. *Nature* **601**, 422–427 (2022).
<https://doi.org/10.1038/s41586-021-04249-w>

- Received: 09 April 2021
- Accepted: 16 November 2021
- Published: 05 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04249-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

Further reading

- [**A blood test to predict complications of pregnancy**](#)

- Lydia L. Shook
- Andrea G. Edlow

Nature (2022)

[**A blood test to predict complications of pregnancy**](#)

- Lydia L. Shook
- Andrea G. Edlow

News & Views 05 Jan 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04249-w>

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

- Article
- [Published: 22 December 2021](#)

Valine tRNA levels and availability regulate complex I assembly in leukaemia

- [Palaniraja Thandapani](#) ORCID: [orcid.org/0000-0002-9881-3645¹](https://orcid.org/0000-0002-9881-3645),
- [Andreas Kloetgen](#) ORCID: [orcid.org/0000-0003-0991-4442^{1,2 na1}](https://orcid.org/0000-0003-0991-4442),
- [Matthew T. Witkowski^{1 na1}](#),
- [Christina Glytsou^{1 na1}](#),
- [Anna K. Lee¹](#),
- [Eric Wang](#) ORCID: [orcid.org/0000-0001-7865-6702¹](https://orcid.org/0000-0001-7865-6702),
- [Jingjing Wang¹](#),
- [Sarah E. LeBoeuf¹](#),
- [Kleopatra Avrampou¹](#),
- [Thales Papagiannakopoulos¹](#),
- [Aristotelis Tsirigos](#) ORCID: [orcid.org/0000-0002-7512-8477^{1,3,4}](https://orcid.org/0000-0002-7512-8477) &
- [Iannis Aifantis](#) ORCID: [orcid.org/0000-0001-6857-1035¹](https://orcid.org/0000-0001-6857-1035)

[Nature](#) volume 601, pages 428–433 (2022)

- 7405 Accesses
- 552 Altmetric
- [Metrics details](#)

Subjects

- [Cancer genomics](#)

- [tRNAs](#)

Abstract

Although deregulation of transfer RNA (tRNA) biogenesis promotes the translation of pro-tumorigenic mRNAs in cancers^{1,2}, the mechanisms and consequences of tRNA deregulation in tumorigenesis are poorly understood. Here we use a CRISPR–Cas9 screen to focus on genes that have been implicated in tRNA biogenesis, and identify a mechanism by which altered valine tRNA biogenesis enhances mitochondrial bioenergetics in T cell acute lymphoblastic leukaemia (T-ALL). Expression of valine aminoacyl tRNA synthetase is transcriptionally upregulated by NOTCH1, a key oncogene in T-ALL, underlining a role for oncogenic transcriptional programs in coordinating tRNA supply and demand. Limiting valine bioavailability through restriction of dietary valine intake disrupted this balance in mice, resulting in decreased leukaemic burden and increased survival *in vivo*. Mechanistically, valine restriction reduced translation rates of mRNAs that encode subunits of mitochondrial complex I, leading to defective assembly of complex I and impaired oxidative phosphorylation. Finally, a genome-wide CRISPR–Cas9 loss-of-function screen in differential valine conditions identified several genes, including *SLC7A5* and *BCL2*, whose genetic ablation or pharmacological inhibition synergized with valine restriction to reduce T-ALL growth. Our findings identify tRNA deregulation as a critical adaptation in the pathogenesis of T-ALL and provide a molecular basis for the use of dietary approaches to target tRNA biogenesis in blood malignancies.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Valine tRNA biogenesis is upregulated by NOTCH1 in T-ALL.

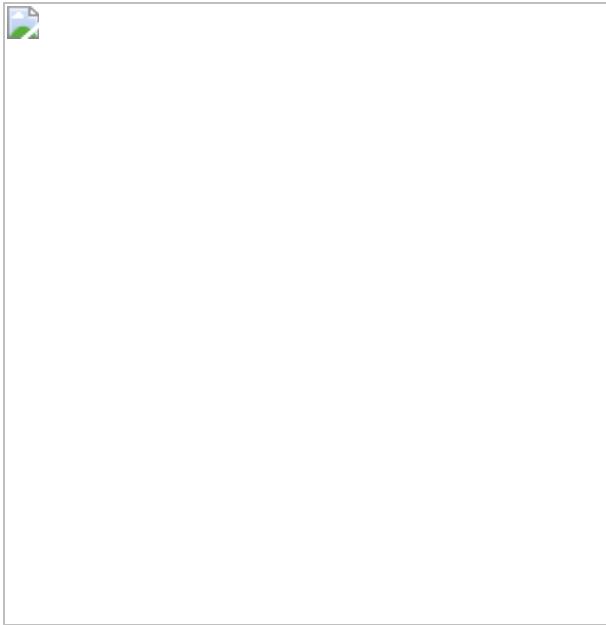


Fig. 2: Dietary valine restriction reduces leukaemic burden and increases survival.

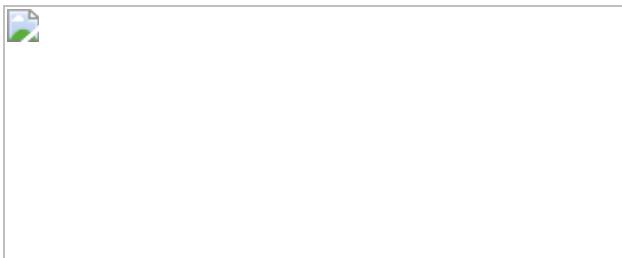
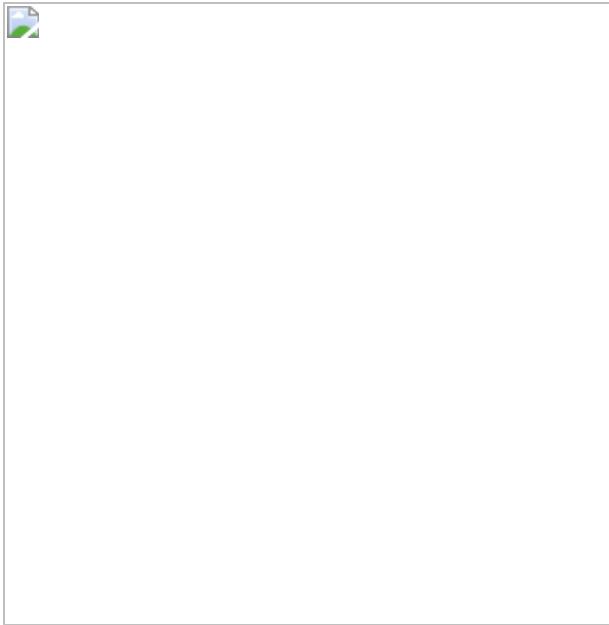


Fig. 3: Dietary valine restriction reduces translation of mRNAs involved in the mitochondrial ETC.



Fig. 4: Valine tRNA biogenesis and bioavailability regulate mitochondrial complex I assembly.



Data availability

All sequencing data created within this study were uploaded to the NCBI Gene Expression Omnibus ([GEO](#)) and is available under the accession codes: [GSE165736](#); [GSE165661](#) for RNA-seq; [GSE165734](#) for tRNA-seq; [GSE167534](#) for the CRISPR screen; and [GSE167535](#) for Ribo-seq. [Source data](#) are provided with this paper.

Code availability

All custom codes written within this study can be found in the [Supplementary Information](#)

References

1. 1.

Rapino, F. et al. Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature* **558**, 605–609 (2018).

2. 2.

Goodarzi, H. et al. Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell* **165**, 1416–1427 (2016).

3. 3.

Wolfe, A. L. et al. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).

4. 4.

Truitt, M. L. et al. Differential requirements for eIF4E dose in normal development and cancer. *Cell* **162**, 59–71 (2015).

5. 5.

Loayza-Puch, F. et al. Tumour-specific proline vulnerability uncovered by differential ribosome codon reading. *Nature* **530**, 490–494 (2016).

6. 6.

Cordo', V., van der Zwet, J. C. G., Canté-Barrett, K., Pieters, R. & Meijerink, J. P. P. T-cell acute lymphoblastic leukemia: a roadmap to targeted therapies. *Blood Cancer Discov.* **2**, 19–31 (2021).

7. 7.

Van Vlierberghe, P. et al. ETV6 mutations in early immature human T cell leukemias. *J. Exp. Med.* **208**, 2571–2579 (2011).

8. 8.

Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).

9. 9.

Gingold, H. et al. A dual program for translation regulation in cellular proliferation and differentiation. *Cell* **158**, 1281–1292 (2014).

10. 10.

Aharon-Hefetz, N. et al. Manipulation of the human tRNA pool reveals distinct tRNA sets that act in cellular proliferation or cell cycle arrest. *eLife* **9**, e58461 (2020).

11. 11.

Weng, A. P. et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).

12. 12.

Palomero, T. et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc. Natl Acad. Sci. USA* **103**, 18261–18266 (2006).

13. 13.

Kloetgen, A. et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat. Genet.* **52**, 388–400 (2020).

14. 14.

Weng, A. P. et al. c-Myc is an important direct target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. *Genes Dev.* **20**, 2096–2109 (2006).

15. 15.

Medyouf, H. et al. High-level IGF1R expression is required for leukemia-initiating cell activity in T-ALL and is supported by Notch signaling. *J. Exp. Med.* **208**, 1809–1822 (2011).

16. 16.

González-García, S. et al. CSL-MAML-dependent Notch1 signaling controls T lineage-specific IL-7Ra gene expression in early human thymopoiesis and leukemia. *J. Exp. Med.* **206**, 779–791 (2009).

17. 17.

Harper, A. E., Miller, R. H. & Block, K. P. Branched-chain amino acid metabolism. *Annu. Rev. Nutr.* **4**, 409–454 (1984).

18. 18.

Aster, J. C. et al. Oncogenic forms of NOTCH1 lacking either the primary binding site for RBP-J κ or nuclear localization sequences retain the ability to associate with RBP-J κ and activate transcription. *J. Biol. Chem.* **272**, 11336–11343 (1997).

19. 19.

King, B. et al. The ubiquitin ligase FBXW7 modulates leukemia-initiating cell activity by regulating MYC stability. *Cell* **153**, 1552–1566 (2013).

20. 20.

Kamata, S. et al. Dietary deprivation of each essential amino acid induces differential systemic adaptive responses in mice. *Mol. Nutr. Food Res.* **58**, 1309–1321 (2014).

21. 21.

Knott, S. R. V. et al. Asparagine bioavailability governs metastasis in a model of breast cancer. *Nature* **554**, 378–381 (2018).

22. 22.

Krall, A. S. et al. Asparagine couples mitochondrial respiration to ATF4 activity and tumor growth. *Cell Metab.* **33**, 1013–1026.e6 (2021).

23. 23.

Taya, Y. et al. Depleting dietary valine permits nonmyeloablative mouse hematopoietic stem cell transplantation. *Science* **354**, 1152–

1155 (2016).

24. 24.

D'Antona, G. et al. Branched-chain amino acid supplementation promotes survival and supports cardiac and skeletal muscle mitochondrial biogenesis in middle-aged mice. *Cell Metab.* **12**, 362–372 (2010).

25. 25.

Cummings, N. E. et al. Restoration of metabolic health by decreased consumption of branched-chain amino acids. *J. Physiol.* **596**, 623–645 (2018).

26. 26.

Solon-Biet, S. M. et al. Branched chain amino acids impact health and lifespan indirectly via amino acid balance and appetite control. *Nat. Metab.* **1**, 532–545 (2019).

27. 27.

Wang, W. & Zou, W. Amino acids and their transporters in T cell immunity and cancer therapy. *Mol. Cell* **80**, 384–395 (2020).

28. 28.

White, P. J. et al. The BCKDH kinase and phosphatase integrate BCAA and lipid metabolism via regulation of ATP-citrate lyase. *Cell Metab.* **27**, 1281–1293.e7 (2018).

29. 29.

Altman, B. J., Stine, Z. E. & Dang, C. V. From Krebs to clinic: glutamine metabolism to cancer therapy. *Nat. Rev. Cancer* **16**, 619–634 (2016).

30. 30.

Nicklin, P. et al. Bidirectional transport of amino acids regulates mTOR and autophagy. *Cell* **136**, 521–534 (2009).

31. 31.

Rosilio, C. et al. L-type amino-acid transporter 1 (LAT1): a therapeutic target supporting growth and survival of T-cell lymphoblastic lymphoma/T-cell acute lymphoblastic leukemia. *Leukemia* **29**, 1253–1266 (2015).

32. 32.

Wempe, M. F. et al. Metabolism and pharmacokinetic studies of JPH203, an L-amino acid transporter 1 (LAT1) selective compound. *Drug Metab. Pharmacokinet.* **27**, 155–161 (2012).

33. 33.

Souers, A. J. et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **19**, 202–208 (2013).

34. 34.

Costa-Mattioli, M. & Walter, P. The integrated stress response: from mechanism to disease. *Science* **368**, eaat5314 (2020).

35. 35.

Giacomello, M., Pyakurel, A., Glytsou, C. & Scorrano, L. The cell biology of mitochondrial membrane dynamics. *Nat. Rev. Mol. Cell Biol.* **21**, 204–224 (2020).

36. 36.

To, T. L. et al. A compendium of genetic modifiers of mitochondrial dysfunction reveals intra-organelle buffering. *Cell* **179**, 1222-1238.e17 (2019).

37. 37.

Herranz, D. et al. Metabolic reprogramming induces resistance to anti-NOTCH1 therapies in T cell acute lymphoblastic leukemia. *Nat. Med.* **21**, 1182–1189 (2015).

38. 38.

Kishton, R. J. et al. AMPK is essential to balance glycolysis and mitochondrial metabolism to control T-ALL cell stress and survival. *Cell Metab.* **23**, 649–662 (2016).

39. 39.

Garcia-Bermudez, J. et al. Aspartate is a limiting metabolite for cancer cell proliferation under hypoxia and in tumours. *Nat. Cell Biol.* **20**, 775–781 (2018).

40. 40.

Manoli, I. & Venditti, C. P. Disorders of branched chain amino acid metabolism. *Transl. Sci. Rare Dis.* **1**, 91–110 (2016).

41. 41.

Haijes, H. A., van Hasselt, P. M., Jans, J. J. M. & Verhoeven-Duif, N. M. Pathophysiology of propionic and methylmalonic acidemias. Part 2: treatment strategies. *J. Inherit. Metab. Dis.* **42**, 745–761 (2019).

42. 42.

Sivanand, S. & Vander Heiden, M. G. Emerging roles for branched-chain amino acid metabolism in cancer. *Cancer Cell* **37**, 147–156 (2020).

43. 43.

Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343**, 80–84

(2014).

44. 44.

Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

45. 45.

Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

46. 46.

Breese, M. R. & Liu, Y. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics* **29**, 494–496 (2013).

47. 47.

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

48. 48.

Pavlova, N. N. et al. Translation in amino-acid-poor environments is limited by tRNAGln charging. *eLife* **9**, e62307 (2020).

49. 49.

Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

50. 50.

Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

51. 51.

Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).

52. 52.

Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).

Acknowledgements

We thank all members of the Aifantis laboratory for discussions throughout this project; S. Tavazoie (Rockefeller University) for sharing the tRNA-seq protocol along with the probe sequences; and N. Sanjana (NYU/NY Genome Center) for sharing the CRISPR–Cas13d tools. P.T. was supported by an AACR Incyte Corporation Leukemia Research Fellowship and by a Young Investigator Grant from Alex’s Lemonade Stand Cancer Research Foundation. M.T.W. is supported by the Leukemia & Lymphoma Society Career Development Program, American Society of Hematology Restart Award, and The Jeffrey Pride Foundation for Pediatric Cancer Research and the Children’s Oncology Group Foundation. C.G. is supported by the NIH/NCI 1K99CA252602-01 grant and is a Special Fellow of the Leukemia & Lymphoma Society. We thank NYU Langone Health Microscopy Laboratory for consultation and assistance with transmission electron microscopy (and light microscopy) work. This shared resource is partially supported by the Cancer Center support grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center. T.P. is supported by NIH grants (R37CA222504 and R01CA227649) and an American Cancer Society research scholar grant (RSG-17-200-01-TBE). A.T. is supported by the grants NC/NIH P01CA229086 and NCI/NIH R01CA252239. We thank the Genome Technology Center (GTC) for sequencing, and the Applied Bioinformatics Laboratories (ABL) for providing bioinformatics support. GTC and ABL are shared resources partially supported by the Cancer Center support grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center. This work used computing resources at the NYU School of

Medicine High Performance Computing Facility. I.A. is supported by the NCI/NIH (1P01CA229086, 1R01CA228135, R01CA216421, R01CA242020, R01CA173636 and 1R01HL159175), the Alex's Lemonade Stand Cancer Research Foundation, the St. Baldrick's Foundation, the Leukemia and Lymphoma Society (TRP no. 6580-20), the Edward P Evans Foundation and the NYSTEM program of the New York State Health Department.

Author information

Author notes

1. These authors contributed equally: Andreas Kloetgen, Matthew T. Witkowski, Christina Glytsou

Affiliations

1. Department of Pathology and Laura & Isaac Perlmutter Cancer Center, NYU School of Medicine, New York, NY, USA

Palaniraja Thandapani, Andreas Kloetgen, Matthew T. Witkowski, Christina Glytsou, Anna K. Lee, Eric Wang, Jingjing Wang, Sarah E. LeBoeuf, Kleopatra Avrampou, Thales Papagiannakopoulos, Aristotelis Tsirigos & Iannis Aifantis

2. Department of Computational Biology of Infection Research, Helmholtz Centre for Infection Research, Braunschweig, Germany

Andreas Kloetgen

3. Applied Bioinformatics Laboratories, NYU School of Medicine, New York, NY, USA

Aristotelis Tsirigos

4. Institute for Computational Medicine, NYU School of Medicine, New York, NY, USA

Aristotelis Tsirigos

Contributions

P.T. conceived, planned and performed most experiments, and co-wrote the manuscript. A.K. performed all the computational analyses with input from A.T. M.T.W. contributed to the conception of using valine-restricted diets and helped with performing the *in vivo* experiments and characterizing overall haematopoiesis following dietary deprivation/restriction of valine. C.G. helped with the mitochondrial experiments and performed the electron micrography sample preparation and analysis. A.K.L. performed the tRNA aminoacylation assays. E.W. helped with the tRNA CRISPR screen. J.W. and K.A. helped with the xenograft experiments. S.E.L. performed the plasma quantification of valine with input from T.P. I.A. directed and coordinated the study, oversaw the results and co-wrote the manuscript. All authors discussed the results and commented on the manuscript.

Corresponding authors

Correspondence to [Palaniraja Thandapani](#) or [Iannis Aifantis](#).

Ethics declarations

Competing interests

I.A. is a consultant for Foresite Labs. A.T. is a scientific advisor to Intelligencia AI. T.P. has received honoraria and/or consulting fees from Calithera Biosciences, Vividion Therapeutics and research support from Bristol Myers Squibb, Dracen Pharmaceuticals and Agios Pharmaceuticals. All other authors declare no competing interests.

Peer review information

Nature thanks Reuven Agami 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 VARS mRNA expression is upregulated in primary T-ALL and is essential in T-ALL.

a, Gene Set Enrichment Analysis (GSEA) of genes involved in tRNA biogenesis in primary T-ALL compared to thymocytes and mature T cells subsets from healthy individuals. **b**, Screen validation of VARS using a competition-based proliferation assay in Cas9 expressing cell lines CUTLL1, Jurkat and DND41. Plotted are GFP percentages measured during 20 days in culture and normalized to day 4. Negative control sgRosa and three independent sgRNAs targeting *VARS* are shown in the graphs ($n=3$; mean \pm SD of biological replicates). **c**, Box plot showing the expression of VARS mRNA in T-ALL patients ($n=57$ individuals) and thymocyte subsets (7 thymocyte and mature T-cell subsets derived from ($n=3$) independent donors) (quantile-normalized microarray results downloaded from GSE33469 and GSE33470). Boxes represent first and third quartiles and the line represent the median. Statistical evaluation between groups was performed using unpaired two-sided t-test followed by multiple testing correction (FDR) (Box plot values for CD34⁺CD1a⁻: min=6.93, max=6.99, q25%=6.94, q50%=6.96, q75%=6.97; for CD34⁺CD1a⁺: min=7.05, max=7.46, q25%=7.07, q50%=7.10, q75%=7.23; for CD4ISP: min=7.79, max=7.92, q25%=7.82, q50%=7.86, q75%=7.89; for DPCD3⁻: min=7.07, max=7.13, q25%=7.09, q50%=7.10, q75%=7.12; for DPCD3⁺: min=6.70, max=7.29, q25%=6.84, q50%=6.97, q75%=7.13; for CD3⁺CD4⁺: min=7.02, max=7.46, q25%=7.14, q50%=7.26, q75%=7.36; for CD3⁺CD8⁺: min=6.91, max=7.22, q25%=6.94, q50%=6.96, q75%=7.09; for T-ALL: min=8.98, max=11.28, q25%=10.29, q50%=10.47, q75%=10.63). **d**, Heatmap representation of changes in gene expression of all the amino acyl tRNA synthetases (RNA-seq) between

primary T-ALL and naïve CD4 T cells. Heatmap shows \log_2 fold-change (FC) of FPKM of T-ALL against average CD4 T cells.

[Source data](#)

Extended Data Fig. 2 tRNA-Val-TAC expression is upregulated in primary T-ALL and is strongly essential for T-ALL survival.

a, Scatter plot showing the total expression of tRNAs in primary T-ALL samples ($n=5$ biological replicates) in the x-axis with differential expression of each tRNA species relative to mature CD4 T cell subset ($n=2$ biological replicates) in y-axis. **b**, Scatter plot showing the total expression of tRNAs in T-ALL cell lines ($n=4$ independent cell lines) in the x-axis with differential expression of each tRNA species relative to mature CD4 T cell subset ($n=2$ biological replicates) in y-axis. **c**, Schematic depicting the CRISPR Cas9 targeting of the anti-codon loop of tRNA-Val-TAC family genes. **d**, Sanger sequencing trace file of tRNA-Val-TAC genomic region in sgROSA and sgVal-TAC transduced CUTLL1 Cas9 cells. **e**, Size of indels along with their frequencies observed as quantified by the Synthego ICE program. **f**, Competition based proliferation assay in Cas9 expressing T-ALL cell lines CUTLL1, Jurkat and DND41. Plotted are GFP^+ percentages measured during 16 days in culture and normalized to day 4. Negative control sgRosa and sgRNA targeting *Val-TAC* tRNA family are shown in the graphs (mean \pm SD of $n=3$ independent experiments).

[Source data](#)

Extended Data Fig. 3 Oncogenic NOTCH1 signaling upregulates valine tRNA biogenesis.

a, GRO-Seq cluster analysis was performed by applying dimensional reduction with a manifold approximation and projection (UMAP) followed by k-means clustering with $k=8$. Clusters 2 ($n = 2105$ genes) and 3 ($n = 2177$ genes) are highlighted in orange and red respectively. **b**, Heatmap representation of mRNA expression changes measured by GRO-Seq within k-means clusters 2 and 3. The clusters define γ SI responding genes with fast

re-establishment of expression post inhibitor wash-off (see Extended Data Fig. 3a for median expression levels of all k-means clusters). Heatmap shows \log_2 FC of FPKM of individual GRO-Seq replicates against average DMSO signal of the respective genes. **c**, Snapshots of H3K4me3 ChIP-Seq on *tRNA-Val-TAC* 1-1 promoter and **d**, *IL7R* promoter as fold-enrichment over input. Below GRO-Seq in CUTLL1 cells upon DMSO, γ SI treatment and release after drug washout as counts-per-million (cpm). **e**, Volcano plot showing differential expression of tRNA pathway genes (153 genes) in CUTLL1 (left) and KOPTK1 (right) upon treatment with DMSO or γ SI for 72h. ($n=2$; FDR < 0.05 and \log_2 FC > 0.58 or < -0.58), cytoplasmic aminoacyl tRNA synthetases are highlighted in green. Statistical evaluation was performed using two-sided edgeR analysis (function glmQLFTest) followed by multiple testing correction (FDR). **f**, Relative expression of VARS mRNA by qPCR analysis in human T-ALL cell line TAL1 following treatment with γ SI for 72h. (mean \pm SD of $n=3$ independent replicates, two-sided unpaired t-test). **g**, Volcano plot of differential expression of tRNA genes in CUTLL1 upon treatment with DMSO or γ SI for 4 days. Valine tRNA genes are highlighted in green ($n=2$; FDR < 0.1 and \log_2 FC > 0.58 or < -0.58). Statistical evaluation was performed using two-sided edgeR analysis (function glmQLFTest) followed by multiple testing correction (FDR). **h**, Val-tRNA and control Leu-tRNA aminoacylation analysis of CUTLL1 cells treated with either DMSO or γ SI (mean \pm SD of $n=3$ independent replicates, two-sided unpaired t-test).

Source data

Extended Data Fig. 4 Dietary valine deprivation/restriction reduce leukemic burden and increases survival in mice.

a, Kaplan-Meier curve representing morbidity of recipient mice transplanted with ckit+ cells transduced with either pMIG ($n=8$) or NOTCH1- Δ E-IRES-GFP ($n=9$) (Log-rank test, two-sided). **b**, Representative image of spleen from mice secondary transplanted with NOTCH- Δ E-GFP+ tumors and fed either control diet or valine-deficient diet for 3 weeks. **c**, Absolute leukemic burden in peripheral blood from mice fed either control diet or valine deficient diet for 3 weeks (mean \pm SD

of $n=5$ animals; two-sided unpaired t-test). **d**, Percentage annexin V positive cells (mean \pm SD of $n=3$ animals; two-sided unpaired t-test). **e**, Absolute leukemic burden in peripheral blood from the mice transplanted with NOTCH- ΔE -GFP $^+$ tumors and fed complete amino acid diet ($n=7$) or diets deficient in either valine ($n=7$), lysine ($n=6$) or asparagine ($n=6$); (mean \pm SD of individual animals; two-sided unpaired t-test). **f**, Kaplan–Meier survival graph of mice represented in (e) (Log-rank test, two-sided). **g**, Representative image of spleen from mice fed different diets in (e) and (f). **h**, Peripheral tumor burden in mice fed valine deficient diet for 2 weeks at late stages of tumor development ($n=10$; two-sided unpaired t-test). **i**, Percentage GFP $^+$ cells in peripheral blood of mice transplanted with (BCR-ABL-GFP $^+$) tumors and fed either control diet or valine-deficient diet for 2 weeks (mean \pm SD of $n = 4$ animals; two-sided unpaired t-test). **j**, Kaplan–Meier survival graph of NSG mice transplanted with patient derived xenograft T-ALL PDX_1 and fed either control diet ($n=5$) or valine deficient diet ($n=5$) (Log-rank test, two-sided). Death owing to leukemia or toxicity of valine deprivation is highlighted. **k**, Body weights of NSG mice transplanted with T-ALL PDX_1 and fed either control valine proficient diet or valine deficient diet (mean \pm SD of $n=5$ animals; two-sided unpaired t-test). **l**, Schematic representation of secondary transplant experiment to test sensitivity of T-ALL to decreasing levels of dietary valine (created with Biorender.com). **m**, Body weights of C57BL/6 mice secondary transplanted with NOTCH- ΔE GFP $^+$ tumors and fed either control valine proficient diet ($n=5$) or valine deficient diet ($n=5$) or valine deficient diet substituted with 0.8g/l valine ($n=4$) or 0.4g/l in drinking water ($n=5$) (mean \pm SD of individual animals; two-sided unpaired t-test).

[Source data](#)

Extended Data Fig. 5 Dietary valine restriction does not affect hematopoiesis.

a, Body weight, White blood cell (WBC), red blood cell (RBC), lymphocyte, hemoglobin (Hgb) and Platelet counts in mice fed either control valine proficient diet (8g/Kg valine), valine deficient diet (0g/Kg valine) or valine deficient diet (0g/Kg valine) substituted with 0.8g/l valine in drinking water (mean \pm SD of $n=5$ animals each condition; two-sided

unpaired t-test) for 4 weeks. **b**, Total cell numbers of bone marrow (tibia+femur), spleen and thymus from mice fed either a control valine proficient diet (8g/Kg valine), valine deficient diet (0g/Kg valine) or valine deficient diet (0g/Kg valine) substituted with 0.8g/l valine in drinking water for 4 weeks (mean ± SD of $n=4$ animals each condition; two-sided unpaired t-test). **c**, Plasma valine levels in peripheral blood serum of mice fed either control valine proficient diet (8g/Kg valine), valine deficient diet (0g/Kg valine) or valine deficient diet (0g/Kg valine) substituted with 0.8g/l valine (mean ± SD of $n=5$ animals each condition; two-sided unpaired t-test).

[Source data](#)

Extended Data Fig. 6 Dietary valine restriction does not affect LSK and thymocytes.

a, Frequency of LSK and different progenitors (GMP, CMP and MEP) from mice fed either a complete valine proficient diet (8g/Kg valine), valine deficient diet (0g/Kg valine) or valine deficient (0g/Kg valine) diet substituted with 0.8g/l valine in drinking water (right) (mean ± SD of $n=4$ animals each condition; two-sided unpaired t-test). **b**, Representative flow cytometry plots of the different hematopoietic compartments highlighted in **(a)**. **c**, Frequency of intra-thymic T-cell population represented as percentage of CD45+ cells from mice fed either a valine proficient diet (8g/Kg valine), valine deficient diet (0g/Kg valine) or valine deficient diet (0g/Kg valine) substituted with 0.8g/l valine in drinking water for 4 weeks (mean ± SD of $n=4$ animals each condition; Two-way Anova analysis with Tukey multiple test correction). **d**, Representative flow cytometry plots of the different thymocyte subsets from **(c)**.

[Source data](#)

Extended Data Fig. 7 Genome-wide CRISPR maps positive and negative genetic interactions with valine restriction.

a, Val-tRNA and control Leu-tRNA aminoacylation analysis of CUTLL1 cells cultured in either standard valine media (20mg/l) or low valine media (2mg/l and 1mg/l) (mean ± SD of $n=3$ independent replicates, two-sided

unpaired t-test). **b**, Gene ranks plotted based on Δ CERES dependency score between low valine media and standard valine media conditions for 4 days. The top negatively and positively selected genes in low valine media are highlighted in blue and red respectively. Genes involved in Branched chain amino acid metabolism are highlighted in green. **c**, Validation of *SLC7A5* dependency and **d**, *BCL2* in low valine media relative to standard valine media using a competition-based survival assay in Cas9 expressing T-ALL cell lines. Plotted are GFP percentages measured during 20 days in culture and normalized to day 4. Negative control sgRosa and two independent sgRNAs targeting *SLC7A5* and *BCL2* are shown in the graphs (mean \pm SD of $n=3$ independent experiments). **e**, IC50 curve of SLC7A5 inhibitor JPH 203 and **f**, BCL2 inhibitor venetoclax of T-ALL cell lines in either standard valine and low valine media. Representative of ($n=2$ of independent experiments).

[Source data](#)

[Extended Data Fig. 8 Valine restriction reduces translation of mitochondrial electron transport chain proteins.](#)

a, Ribosome protected fragment (RPF) length distribution of the Ribo-Seq datasets ($n=2$ biological replicates each condition). **b**, Relative CDS distribution of the RPF fragments ($n=2$ biological replicates each condition). **c**, Percentage valine (left) and lysine (right) codons between mRNAs with lowest translational efficiency (TE; bottom 5%) and mRNAs with no significant changes in translational efficiency ($\log_{2}FC > -0.1375$ and < 0.1375) between leukemic cells isolated from mice fed control diet and valine restricted diet ($n=2$ biological replicates each condition) (Box plot values for valine: for genome 0.8g/l: min=1.42, max=13.73, q25%=4.99, q50%=5.99, q75%=7.09; for bottom 5% 0.8g/l: min=0.00, max=11.84, q25%=5.03, q50%=6.10, q75%=7.42; for genome 0.4g/l: min=0.00, max=17.07, q25%=5.04, q50%=6.09, q75%=7.20; for bottom 5% 0.4g/l: min=0.00, max=11.75, q25%=5.21, q50%=6.18, q75%=7.39; for genome 0g/l: min=0.00, max=15.79, q25%=5.02, q50%=6.9, q75%=7.19; for bottom 5% 0g/l: min=0.00, max=17.07, q25%=5.08, q50%=6.29, q75%=7.36) (Box plot values for lysine: for genome 0.8g/l: min=0.00, max=29.02, q25%=4.28, q50%=5.75, q75%=7.41; for bottom 5% 0.8g/l:

min=0.00, max=14.85, q25%=3.77, q50%=5.36, q75%=7.14; for genome 0.4g/l: min=0.34, max=27.93, q25%=4.36, q50%=5.87, q75%=7.43; for bottom 5% 0.4g/l: min=0.47, max=14.85, q25%=3.92, q50%=5.45, q75%=7.11; for genome 0g/l: min=0.47, max=27.92, q25%=4.43, q50%=6.05, q75%=7.67; for bottom 5% 0g/l: min=0.00, max=15.76, q25%=3.63, q50%=5.26, q75%=6.96). **d**, Volcano plots of Δ translational efficiency in NOTCH- Δ E-GFP⁺ cells isolated from mice fed different levels of dietary valine ($n=2$ of biological replicates each condition). (p -value < 0.05 and $\log_2\text{FC} > 0.58$ or < -0.58). **e**, STRING Network Pathway enrichment analysis of the 162 mRNAs with reduced translational efficiencies in mice fed valine restricted diet relative to mice fed control valine proficient diet. **f**, Snapshot of Ribosome protected fragment (RPF) and total RNA tracks for gene *Ndufs7* and **g**, *Ndufs3*. **h**, RPKM values of total RNA and RPF read counts for *Ndufs3* and *Ndufs7* ($n=2$ of biological replicates each condition). **i**, Schematic of the reporter constructs expressing GFPd2 in fusion with either NDUFB1 (high valine content) or NDUFS5 (low valine content). mCherry serves as an internal control. **j**, CUTLL1 and Jurkat cells infected with the reporter constructs were cultured in either complete RPMI media or media lacking either valine or tryptophan for 8h. Quantification of GFPd2 fluorescence normalized to mCherry signal and cycloheximide treatment are plotted. (mean \pm SD of $n=4$ independent replicates; two-sided unpaired t-test.). **k**, Gene signatures upregulated in control diet (8g/Kg valine) is shown in red whereas gene signatures upregulated in mice fed reduced valine is highlighted in green. **l**, Immunoblots for ATF4 and actin (same gel) from NOTCH- Δ E-GFP⁺ cellular lysates isolated from mice fed different levels of dietary valine ($n=2$ biological replicates). For gel source data, see Supplementary Fig. 1. RPKM, reads per kilobase transcript per million mapped reads.

Source data

Extended Data Fig. 9 Valine tRNA biogenesis and bioavailability regulates levels of mitochondrial complex I.

a, Blue Native (BN)-PAGE analysis (same gel, for gel source data, see Supplementary Fig. 1) and **b**, maximal respiration of KOPTK1 cells cultured in different levels of valine for 72 h. **c**, Oxygen consumption rate

(OCR) and **d**, Maximal respiration of CUTLL1 cells cultured in different levels of valine for 72 h (mean \pm SD of $n=5$ independent replicates; two-sided unpaired t-test). **e**, Relative expression of VARS mRNA by qPCR analysis following 4 days of doxycycline treatment to induce Cas13d expression. ($n=2$ independent replicates). **f**, Competition based proliferation assay to validate loss of fitness following VARS knockdown in CUTLL1 Cas13d cells. Plotted are GFP percentages measured during 20 days in culture and normalized to day 0 of doxycycline treatment. Negative control sgRosa and three independent sgRNAs targeting VARS are shown in the graphs (mean \pm SD of $n=3$ independent replicates). **g**, BN-PAGE (top) analysis of CUTLL1 cells expressing inducible Cas13d and sgRNA_1 targeting VARS transcript. SDS-PAGE and immunoblots confirming VARS knockdown (same gel, for gel source data, see Supplementary Fig. 1). **h**, NAD $^+$ /NADH ratio of CUTLL1 cells cultured in RPMI with different levels of valine for 72 h. Representative of ($n=2$) independent experiments. **i**, MitoSox staining of CUTLL1, KOPTK1 cells cultured in RPMI with different levels of valine for 72 h and analyzed by flow cytometry (left, middle). CUTLL1 Cas13d cells transduced with three independent sgRNAs targeting VARS transcript was treated with doxycycline to induce Cas13 expression. Day 8 of doxycycline treatment, cells were stained with MitoSOX and analyzed by flow cytometry (right). Graphs show Mito-SOX mean fluorescence intensity (MFI). Representative of ($n=3$) independent experiments (mean \pm SD of $n=4$ technical replicates). **j**, Electron micrograph from CUTLL1 cells illustrating the mitochondrial cristae structure. Arrow denotes the parameter of maximal Cristae Lumen Width (CLW) quantified in this study. **k**, Representative electron micrographs of CUTLL1 cells cultured in RPMI with different levels of valine for 72 h. **l**, Representative electron micrographs of CUTLL1-Cas13d cells transduced with either sgROSA or sgRNA targeting VARS and doxycycline treated for 8 days to induce Cas13d expression (left). Quantification of maximal cristae width in 10 randomly selected mitochondria from 15 cells (right) ($n=110$ cristae; mean \pm SD; two-sided unpaired t-test).

[Source data](#)

[Extended Data Fig. 10 Loss of pathway intrinsic ETC proteins promotes resistance to valine restriction.](#)

a, Validation of NDUFB4 as a positively selected gene in low valine media relative to standard valine media using a competition-based proliferation assay in Cas9 expressing cell lines CUTLL1 and DND41. Plotted are GFP percentages measured during 20 days in culture and normalized to day 4. Negative control sgRosa and two independent sgRNAs targeting NDUFB4 are shown in the graphs (mean \pm SD of $n=3$ independent replicates). **b**, Schematic depicting how upregulated valine tRNA biogenesis regulates mitochondrial bio-energetics in T-ALL (created with Biorender.com).

[Source data](#)

Supplementary information

[Supplementary Information](#)

This file contains Supplementary Fig. 1 (uncropped gel images), Supplementary Figs. 2–9 (FACS plots gating strategy) and Supplementary Custom Codes.

[Reporting Summary](#)

[Supplementary Table 1](#)

Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

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

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

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

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

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

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

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

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

[Source Data Extended Data Fig. 9](#)

[Source Data Extended Data Fig. 10](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Thandapani, P., Kloetgen, A., Witkowski, M.T. *et al.* Valine tRNA levels and availability regulate complex I assembly in leukaemia. *Nature* **601**, 428–433 (2022). <https://doi.org/10.1038/s41586-021-04244-1>

- Received: 02 March 2021
- Accepted: 15 November 2021
- Published: 22 December 2021

- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04244-1>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04244-1>

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

- Article
- Open Access
- [Published: 22 December 2021](#)

Targeting SWI/SNF ATPases in enhancer-addicted prostate cancer

- [Lanbo Xiao^{1,2}](#)^{nal},
- [Abhijit Parolia^{1,2,3}](#)^{nal},
- [Yuanyuan Qiao](#) [ORCID: orcid.org/0000-0002-1178-3480](#)^{1,2,4}^{nal},
- [Pushpinder Bawa](#) [ORCID: orcid.org/0000-0002-8370-3638](#)^{1,2},
- [Sanjana Eyunni^{1,2,3}](#),
- [Rahul Mannan](#) [ORCID: orcid.org/0000-0002-6642-0468](#)^{1,2},
- [Sandra E. Carson](#) [ORCID: orcid.org/0000-0002-7716-9627](#)^{1,2},
- [Yu Chang^{1,2}](#),
- [Xiaoju Wang^{1,2,4}](#),
- [Yuping Zhang^{1,2}](#),
- [Josh N. Vo^{1,2,5}](#),
- [Steven Kregel^{1,2}](#),
- [Stephanie A. Simko^{1,2}](#),
- [Andrew D. Delekta^{1,2}](#),
- [Mustapha Jaber¹](#),
- [Heng Zheng^{1,2}](#),
- [Ingrid J. Apel^{1,2}](#),
- [Lisa McMurry^{1,2}](#),
- [Fengyun Su^{1,2}](#),
- [Rui Wang^{1,2}](#),
- [Sylvia Zelenka-Wang^{1,2}](#),
- [Sanjita Sasmal⁶](#),
- [Leena Khare](#) [ORCID: orcid.org/0000-0001-9860-7887](#)⁶,
- [Subhendu Mukherjee⁶](#),
- [Chandrasekhar Abbineni⁶](#),
- [Kiran Aithal⁶](#),
- [Mital S. Bhakta⁷](#),
- [Jay Ghurye⁷](#),

- [Xuhong Cao](#)^{1,2,8},
- [Nora M. Navone](#) ORCID: [orcid.org/0000-0001-8645-6890](#)⁹,
- [Alexey I. Nesvizhskii](#) ORCID: [orcid.org/0000-0002-2806-7819](#)^{1,2,4,5},
- [Rohit Mehra](#) ORCID: [orcid.org/0000-0002-6955-8884](#)^{1,2,4},
- [Ulka Vaishampayan](#)¹⁰,
- [Marco Blanchette](#)⁷,
- [Yuzhuo Wang](#) ORCID: [orcid.org/0000-0002-9749-8591](#)^{11,12},
- [Susanta Samajdar](#)⁶,
- [Murali Ramachandra](#)⁶ &
- [Arul M. Chinnaiyan](#) ORCID: [orcid.org/0000-0001-9282-3415](#)^{1,2,4,5,8,13}

[Nature](#) volume 601, pages 434–439 (2022)

- 22k Accesses
- 246 Altmetric
- [Metrics details](#)

Subjects

- [Chromatin remodelling](#)
- [Prostate cancer](#)
- [Targeted therapies](#)

Abstract

The switch/sucrose non-fermentable (SWI/SNF) complex has a crucial role in chromatin remodelling¹ and is altered in over 20% of cancers^{2,3}. Here we developed a proteolysis-targeting chimera (PROTAC) degrader of the SWI/SNF ATPase subunits, SMARCA2 and SMARCA4, called AU-15330. Androgen receptor (AR)⁺ forkhead box A1 (FOXA1)⁺ prostate cancer cells are exquisitely sensitive to dual SMARCA2 and SMARCA4 degradation relative to normal and other cancer cell lines. SWI/SNF ATPase degradation rapidly compacts *cis*-regulatory elements bound by transcription factors that drive prostate cancer cell proliferation, namely AR, FOXA1, ERG and MYC, which dislodges them from chromatin, disables their core enhancer circuitry, and abolishes the downstream oncogenic gene programs. SWI/SNF ATPase degradation also disrupts super-enhancer and promoter looping interactions that wire supra-physiologic expression of the *AR*, *FOXA1* and *MYC* oncogenes themselves. AU-15330 induces potent inhibition of tumour growth in xenograft models of prostate cancer and synergizes with the AR antagonist enzalutamide, even inducing disease

remission in castration-resistant prostate cancer (CRPC) models without toxicity. Thus, impeding SWI/SNF-mediated enhancer accessibility represents a promising therapeutic approach for enhancer-addicted cancers.

[Download PDF](#)

Main

In eukaryotic cells, DNA is wrapped around histone octamers (referred to as nucleosomes), which form a physical barrier to DNA-based processes⁴. Thus, gene expression is regulated by modifying physical accessibility of the DNA through nucleosomal remodelling and, when in an accessible state, through binding of transcription factors^{5,6}. In this regulatory context, non-coding genomic elements called enhancers have emerged as central hubs serving as integrative platforms for transcription factor binding and activation of lineage-specific gene programs^{7,8}. The enhancer elements can lie within untranslated or distal intergenic regions and make looping interactions with their target gene promoters to potentiate RNA polymerase II (PolII)-mediated transcription^{9,10}.

In cancer, genetic alterations invariably lead to an aberrant transcriptional state that is often wired through expansion and remodelling of the enhancer landscape^{11,12}. This includes de novo commissioning of new enhancers (neo-enhancers) by reprogramming of pioneer factor cistromes¹³, enhancer hijacking via structural rearrangements^{14,15}, and/or abnormal enhancer–promoter interactions via alterations in chromatin topology¹⁶—all to enable hyper-expression of driver oncogenes. Although there has been intense interest in therapeutically targeting aberrant enhancer function in cancer, the molecular machinery responsible for enhancer maintenance and/or activation remains poorly characterized.

Recent studies have uncovered alterations in genes encoding constituent subunits of the SWI/SNF complex in over 20% of human cancers². SWI/SNF is a multi-subunit chromatin-remodelling complex that uses energy from ATP hydrolysis to reposition or eject nucleosomes at non-coding regulatory elements, thereby enabling free DNA access for the transcriptional machinery¹. In SWI/SNF-mutant tumours, the residual complex is thought to enable oncogenic transcriptional programs and speculated to be a viable therapeutic target^{17,18,19}. Although inhibitors and degraders of ATPase and BRD7–BRD9 SWI/SNF subunits have been recently developed^{20,21,22}, to our knowledge, no studies have comprehensively assessed the therapeutic efficacy of SWI/SNF inactivation across a wide spectrum of cancers. To this end, we have developed and characterized a highly-selective PROTAC degrader of both SWI/SNF

ATPase subunits—SMARCA2 (BRM) and SMARCA4 (BRG1)—that are required for the nucleosomal-remodelling functions of SWI/SNF complexes.

We found enhancer-binding transcription factor-addicted cancers (for example, AR-FOXA1-driven prostate cancer) to be exquisitely and preferentially sensitive to SWI/SNF ATPase degradation, which triggered an instantaneous, specific loss of physical accessibility and transcription factor binding at enhancer elements, thereby disrupting enhancer-wired oncogenic gene programs. To our knowledge, this study is the first preclinical proof of concept that targeted obstruction of chromatin accessibility at enhancer elements may be a potent therapeutic strategy in transcription factor-addicted tumours.

Results

We developed the PROTAC degrader, AU-15330, comprising a bait moiety that binds the bromodomain in SMARCA2 and SMARCA4 and a ligand moiety for the von Hippel–Lindau (VHL) ubiquitin ligase (Fig. [1a](#), Extended Data Fig. [1a](#)). AU-15330 also binds to the secondary SWI/SNF module component PBRM1, which relies on the ATPase module for assembly onto the core complex^{[23](#)}. Although it binds to the same bromodomain in target proteins as the PROTAC degrader ACBI1^{[20](#)}, AU-15330 comprises a distinct linker structure that largely dictates a PROTAC’s target selectivity and degradation kinetics^{[24](#)}. Treatment of several cell lines with AU-15330 led to time and dose-dependent degradation of SMARCA2, SMARCA4 and PBRM1 (Fig. [1b](#)). Mass spectrometry-based proteomics analysis confirmed SMARCA2, SMARCA4 and PBRM1 as the only significantly downregulated proteins (Extended Data Fig. [1b](#)). Of note, we detected no change in the abundance of other bromodomain-containing proteins or non-targeted SWI/SNF subunits (Extended Data Fig. [1c, d](#)). SWI/SNF complexes have been shown to assemble in a modular manner, with the ATPase module being the last to bind to the SMARCC1 (also known as BAF155)-containing core complex^{[23](#)}. Accordingly, SMARCC1 nuclear immunoprecipitation followed by mass spectrometry showed no changes in the sequential assembly of the core and secondary modules but revealed detachment of ATPase module subunits upon AU-15330 treatment (Extended Data Fig. [1e](#)).

Fig. 1: AU-15330, a specific degrader of SWI/SNF ATPases, exhibits preferential cytotoxicity in enhancer-binding transcription factor-driven cancers.

 **figure 1**

a, Structure of AU-15330 and schematic of SMARCA2, SMARCA4 and PBRM1 domains. AU-15330-targeted bromodomains (BD) are shown. QLQ, conserved Gln, Leu, Gln motif containing domain; HSA, helicase/SANT-associated domain; BRK, Brahma and Kismet domain; SnAC, Snf2 ATP coupling domain; BAH1, bromo-adjacent homology domain 1; BAH2, bromo-adjacent homology domain 2. **b**, Immunoblots of SMARCA2, SMARCA4 and PBRM1 on treatment of HEK 293 and HeLa cells with AU-15330 at increasing concentrations or time durations. Vinculin is used as a loading control, and is probed on a representative immunoblot. This experiment was repeated independently twice. **c**, IC₅₀ of AU-15330 in a panel of human-derived cancer or normal cell lines after 5 days of treatment. Known SMARCA4 loss-of-function (LOF) alterations and multiple myeloma (MM) cell lines with MYC rearrangements (MYC-R'ed) are identified below the graph. AR and FOXA1 scores quantify their transcriptional activities using cognate multi-gene signatures.

[Source data](#)

Using a panel of normal and cancer cell lines from 14 distinct lineages, we found AR and FOXA1-driven prostate cancer cells to be preferentially sensitive to AU-15330 (half-maximal inhibitory concentrations (IC₅₀) < 100 nM; Fig. [1c](#), Extended Data Fig. [1f,g](#), Supplementary Table [1](#)). AR⁻FOXA1⁻ prostate cancer cells showed moderate sensitivity (IC₅₀ between 100–400 nM), whereas normal and non-neoplastic prostate cells were resistant (IC₅₀ > 1,000 nM) to AU-15330. We observed a similar cytotoxicity profile for ACBI1 and BRM014, an allosteric dual inhibitor of SMARCA2 and SMARCA4 ATPase activity²⁵ (Extended Data Fig. [1h,i](#)). Notably, AR⁺FOXA1⁺ prostate cancer cells were more sensitive to these inhibitors than SMARCA4-null cancer cell lines. Several MYC-driven multiple myeloma cells and

oestrogen receptor- and/or AR-positive breast cancer cells were also acutely sensitive to AU-15330 (Fig. 1c, Extended Data Fig. 1j,k).

In several prostate cancer cell lines, we detected substantial expression of both SWI/SNF ATPases, which were rapidly degraded in a dose-dependent manner by AU-15330 (Extended Data Fig. 2a,b). Concordantly, AU-15330 attenuated the growth of these cells and induced apoptotic cell death, while having no anti-proliferative effect on benign or non-neoplastic prostate cells (grey bars, Fig. 1c) at parallel doses (Extended Data Fig. 1f, 2c–e). Treatment with either the bromodomain ligand alone (AU-15139) or an inactive epimer of AU-15330 (AU-16235) had no effect on target protein levels or cancer cell survival and growth (Extended Data Figs. 1f, g, 2f,g). Next, competition of AU-15330 with a free VHL ligand (VL285), but not with thalidomide, reversed degradation of SWI/SNF targets (Extended Data Fig. 2g) and rescued the growth inhibitory effect in a dose-dependent manner (Extended Data Fig. 2h). Furthermore, pre-treatment of VCaP cells (an AR⁺FOXA1⁺ prostate cancer cell line model) with bortezomib (a proteasome inhibitor) or MLN4924 (a NEDD8-activating enzyme inhibitor) hindered target protein degradation, indicating that AU-15330 requires the proteasome machinery and ubiquitination cascade for its action (Extended Data Fig. 2g).

As SWI/SNF complexes actively remodel nucleosomal DNA packaging, we profiled the effect of AU-15330 on physical chromatin accessibility using the assay for transposase-accessible chromatin followed by sequencing (ATAC-seq). We detected a rapid and near-complete loss in chromatin accessibility at more than 30,000 sites in VCaP cells with as little as 1 h of AU-15330 treatment (Fig. 2a), which is within minutes of SMARCA2 and SMARCA4 degradation (Extended Data Fig. 3a); approximately 25,000 genomic sites showed little to no change in nucleosomal density (Extended Data Fig. 3b). Similar profound changes in chromatin accessibility were not observed upon treatment with a BRD4 degrader (ZBC-260; Extended Data Fig. 3a,b). In our genetic models using CRISPR–Cas9 and shRNA-mediated target inactivation, we detected a significant compaction of the chromatin only upon concurrent loss of both SWI/SNF ATPases (Extended Data Fig. 3c,d). More than 90% of the AU-15330-compacted sites were within distal regulatory regions, which were enriched for enhancers, whereas the retained sites were predominantly within promoters (Fig. 2b). De novo motif and binding analysis for the regulation of transcription (BART) analyses of AU-15330-compacted sites identified DNA-binding elements for major oncogenic transcription factors in prostate cancer, including AR, FOXA1, HOXB13 and ERG (Extended Data Fig. 3e,f). As expected, retained promoter sites showed enrichment for PolII and E2F motifs (Extended Data Fig. 3g). Interrogation of chromatin changes in LNCaP cells upon AU-15330 treatment reproduced these findings (Extended Data Fig. 4a–c).

Fig. 2: SWI/SNF ATPase degradation disrupts physical chromatin accessibility at the core-enhancer circuitry to disable oncogenic transcriptional programs.

 figure 2

a, ATAC-seq read-density heat maps from VCaP cells treated with DMSO or AU-15330 for indicated durations ($n = 2$ biological replicates). **b**, Genome-wide changes in chromatin accessibility upon AU-15330 treatment for 4 h in VCaP cells along with genomic annotation of sites that lose physical accessibility (lost) or remain unaltered (retained). **c, d**, ChIP-seq read-density heat maps for AR and FOXA1 (**c**) and

H3K27Ac (**d**) at the AU-15330 (AU)-compacted genomic sites in VCaP cells after treatment with DMSO or AU-15330 (1 μ M) for indicated times and stimulation with R1881 (1 nM, 3 h). **e**, RNA-seq heat maps for classical AR target genes in LNCaP, VCaP and LAPC4 prostate cancer cells with or without 24 h of AU-15330 treatment.

Source data

Concurrent with the loss of accessibility, chromatin immunoprecipitation followed by sequencing (ChIP-seq) revealed a decrease in chromatin binding of AR, FOXA1, and ERG in VCaP cells within 1 h of AU-15330 treatment (Fig. [2c](#), Extended Data Fig. [4d](#), [e](#)). We also detected disappearance of the characteristic ‘valley’ pattern in the H3K27Ac ChIP-seq signal, indicating the movement of flanking nucleosomes towards the centre of AU-15330-compacted enhancers (Fig. [2d](#)). At early time points, we detected no loss in the abundance of the H3K27Ac mark; however, it was significantly depleted 24 h after AU-15330 treatment (Extended Data Fig. [4f](#)). Similar results were observed upon AU-15330 treatment of LNCaP cells (Extended Data Fig. [4g](#), [h](#)). Loss of AR, FOXA1 and H3K27Ac ChIP signals was evident at enhancer sites of the classical AR target gene *KLK3* (Extended Data Fig. [4i](#)). We found AR, FOXA1, ERG and SMARCC1 to co-occupy a large fraction of H3K27Ac-marked regulatory elements (Extended Data Fig. [5a–c](#)). Furthermore, multiple core SWI/SNF components were present in the mass spectrometry-based datasets of AR, FOXA1, and ERG interactomes (Extended Data Fig. [5d](#)), which we confirmed by reciprocal co-immunoprecipitation assays (Extended Data Fig. [5e](#)). This positions SWI/SNF complexes as common chromatin cofactors of the oncogenic transcriptional machinery in prostate cancer cells. As an important control, we saw no changes in chromatin binding of CTCF in AU-15330-treated cells (Extended Data Fig. [6a–d](#)).

Global transcriptomic profiling with RNA sequencing (RNA-seq) revealed significant downregulation of AR and FOXA1-regulated genes in multiple prostate cancer cells, as well as ERG-regulated transcripts in ERG fusion-positive VCaP cells. We also detected significant loss in the expression of MYC target genes with AU-15330 (Fig. [2e](#), Extended Data Fig. [6e](#), [f](#)). The global AU-15330 gene signature was highly concordant with transcriptional changes associated with ARID1A loss (Extended Data Fig. [6g](#)). However, neither BRD7 nor BRD9 degradation alone attenuated the expression of classical AR, FOXA1 and ERG target genes or the *MYC* gene to an extent comparable to AU-15330, suggesting that canonical SWI/SNF (cBAF) complexes are the primary cofactors of oncogenic enhancer-binding transcription factors (Extended Data Fig. [6h–j](#)). The expressions of *AR*, *MYC* and *FOXA1* genes themselves are frequently amplified in advanced prostate cancer by copy amplification and/or enhancer duplication^{[15,26,27](#)}. We found that AU-15330 markedly decreased expression of *AR*, *FOXA1*, *MYC* and *TMPRSS2–ERG* transcripts to 40–60% of their baseline expression (Extended Data Fig. [7a](#)), with parallel decreases at the protein level (Fig. [3a](#)). More severe transcriptional attenuation of these oncogenes was noted

upon BRD4 degradation by ZBC-260, with AU-15330 specifically abolishing expression of additional driver oncogenes (Extended Data Fig. [7b](#)), again suggesting a distinct mechanism of action for AU-15330-mediated anti-tumour cytotoxicity. Similar results were observed in genetic-inactivation models (Extended Data Fig. [7c](#)).

Fig. 3: SWI/SNF ATPase degradation disrupts enhancer–promoter loops to temper supra-physiologic expression of driver oncogenes.

 **figure 3**

a, Immunoblots of indicated proteins in VCaP cells treated with DMSO for 24 h or AU-15330 (1 μ M) for increasing time durations. GAPDH is used as a loading control, and is probed on a representative immunoblot. This experiment was repeated independently twice. **b**, H3K27Ac ChIP-seq signal rank-ordered list of super-

enhancers in VCaP cells with select *cis*-coded driver oncogenes denoted (HOMER). **c**, Normalized read density of ATAC-seq at super-enhancers ($n = 32,545$ sites) in VCaP cells treated with DMSO or AU-15330 (1 μ M) for 1 or 4 h (two-sided t-test). In box plots, the centre line shows median, box edges mark quartiles 1–3, and whiskers span quartiles $1–3 \pm 1.5 \times$ interquartile range. **d**, H3K4me3 HiChIP-seq heat maps within the *AR* gene locus in VCaP cells with or without AU-15330 (1 μ M) treatment for 4 h (bin size = 25 kb). ATAC-seq read-density tracks from the same treatment conditions are overlaid. Grey highlights mark enhancers; blue highlights the *AR* promoter. Loops indicate read-supported *cis* interactions within the locus. IR, interaction reads. **e**, APA plots for H3K4me3 and H3K27Ac HiChIP-seq data for all possible interactions between putative enhancers and gene promoters in VCaP cells with or without with AU-15330 treatment (1 μ M, 4 h).

Source data

The hyper-expression of oncogenes like *AR*, *FOXA1* and *MYC* in cancer has been shown to be wired through looping interactions with multi-enhancer clusters^{15,26,27}, often referred to as super-enhancers. Several such regulatory clusters were identified in *cis*-proximity of the *AR*, *MYC* and *TMPRSS2-ERG* genes (Fig. 3b), and AU-15330 treatment led to immediate compaction of these sites and loss of H3K27Ac, AR and FOXA1 ChIP-seq signal at the super-enhancers (Fig. 3c, Extended Data Fig. 7d). To detect changes in the interaction of super-enhancers with their target gene promoters, we performed H3K4me3 (active promoter mark) and H3K27Ac Hi-C coupled with ChIP-seq (HiChIP-seq) upon AU-15330 treatment. SWI/SNF inactivation markedly disrupted the three-dimensional looping interactions of *cis*-enhancers with the *AR* gene promoter (Fig. 3d, Extended Data Fig. 8a). Similar attenuation of enhancer–promoter interactions was detected by H3K27Ac HiChIP-seq at the *FOXA1* locus (Extended Data Fig. 8b), which is recurrently rearranged in advanced prostate cancer¹⁵. Aggregate peak analyses (APA) of enhancer–promoter interactions showed a marked attenuation of contact strength and/or frequency starting as early as 2 h after AU-15330 treatment, that is, within 1 h of SMARCA2 and SMARCA4 degradation (Fig. 3e, Extended Data Fig. 8c). At these early time points, we did not detect a significant decrease in H3K27Ac signal at the compacted enhancer sites (Fig. 2d), strongly suggesting that physical chromatin accessibility and transcription factor binding serve as primary determinants of functional enhancer–promoter interactions. Of note, we found no change in the looping interactions between CTCF-bound elements (Extended Data Fig. 8d,e). Together, these data show that SWI/SNF ATPase inactivation specifically leads to genome-wide collapse of the AR, FOXA1, ERG and MYC-activated core enhancer circuitry in prostate cancer cells.

Next, we pharmacologically characterized AU-15330 in animal models of advanced prostate cancer. Notably, prolonged AU-15330 treatments showed no evident toxicity

in immuno-competent mice (Extended Data Fig. 9, [Supplementary Text](#), Supplementary Table 3). We first employed the VCaP castration-resistant prostate cancer (CRPC) model (VCaP-CRPC) to assess the efficacy of AU-15330. As expected, treatment of castrated male mice bearing the VCaP-CRPC xenografts with enzalutamide (an AR antagonist) showed moderate anti-tumour efficacy; however, treatment with AU-15330 led to potent inhibition of tumour growth, triggering disease regression in more than 20% of animals (Fig. 4a, b, Extended Data Fig. 10a, b). Treatment with the combinatorial regimen (AU-15330 plus enzalutamide) induced the most potent anti-tumour effect, with regression in all animals (Fig. 4a, b, Extended Data Fig. 10b). Tumours showed robust downregulation of SWI/SNF targets and AR, ERG, MYC and Ki67 after five days of AU-15330 treatment, both when administered alone or with enzalutamide (Fig. 4c, Extended Data Fig. 10c–e). No significant change in body weight was noted throughout any of these treatments, nor was there any histologic evidence of toxicity in essential organs at endpoint (Extended Data Fig. 10f–h). AU-15330 also strongly inhibited the growth of C4-2B cell line-derived CRPC xenografts in intact mice as a single agent and synergized with enzalutamide (Fig. 4d, Extended Data Fig. 11a–d). An in vitro evaluation of drug synergism between AU-15330 and enzalutamide confirmed synergism of the two drugs in multiple prostate cancer cell lines (Extended Data Fig. 11e–h), and pre-treatment with either drug significantly reduced the IC₅₀ value of the other (Extended Data Fig. 11i, j).

Fig. 4: AU-15330 inhibits tumour growth in preclinical models of CRPC and synergizes with enzalutamide.

 **figure 4**

a, Tumour volume (measured twice per week using callipers) in the VCaP-CRPC model with AU-15330 alone or in combination with enzalutamide (two-sided *t*-test). Data are mean \pm s.e.m. (vehicle: $n = 18$; AU-15330: $n = 20$; enzalutamide: $n = 18$; AU-15330 + enzalutamide: $n = 16$). **b**, Waterfall plot depicting change in tumour volume after 33 days of treatment. Response evaluation criteria in solid tumours (RECIST) was used to stratify tumours: progressive disease (PD), at least a 20% increase in tumour size; stable disease (SD), increase of <20% to a decrease of <30%; partial response (PR), at least a 30% decrease. The vehicle and enzalutamide groups have 100% PD; the AU-15330 group has 61% PD, 33% SD and 6% PR; and the AU-15330 + enzalutamide group has 0% PD, 12% SD and 88% PR. **c**, Representative haematoxylin and eosin (H&E) staining and immunohistochemistry from the VCaP-CRPC xenograft study ($n = 2$ tumours per condition). Insets in the H&E images show expanded views of apoptotic cells. **d**, Tumour volume measurements showing efficacy of AU-15330, enzalutamide or combined treatment in C4-2B-derived CRPC

xenografts ($n = 20$ per condition; two-sided t -test). Data are mean \pm s.e.m. **e**, Tumour volume measurements showing the effect of enzalutamide alone or in combination with AU-15330 in the castration-resistant MDA-PCa-146-12 PDX study (two-sided t -test). Data are mean \pm s.e.m. **f**, Mechanism of action of AU-15330-triggered cytotoxicity in AR–FOXA1-signalling-driven prostate cancer. SWI/SNF ATPase degradation induces a rapid, targeted loss in chromatin accessibility at the core-enhancer circuitry of AR, FOXA1, ERG and MYC, thereby attenuating their cancer-promoting transcriptional programs and tempering the enhancer-wired supra-physiologic expression of driver oncogenes.

Source data

Treatment with AU-15330 was similarly effective in inhibiting the growth of enzalutamide-resistant cell lines, including derivatives of VCaP and LNCaP cells (Extended Data Fig. [11k–l](#)). The combinatorial regimen also markedly inhibited tumour growth in MDA-PCa-146-12, a patient-derived xenograft (PDX) model that is inherently resistant to enzalutamide (Extended Data Fig. [12a–c](#)). We further established a CRPC variant of the MDA-PCa-146-12 PDX by tumour implantation into castrated mice (Extended Data Fig. [12a](#)). Even in this highly aggressive model, the combinatorial regimen induced significant tumour growth inhibition, causing regression in more than 30% of animals (Fig. [4e](#), Extended Data Fig. [12d](#)). In all arms of these studies, we detected no changes in animal body weights (Extended Data Fig. [12e,f](#)). There was also no sign of goblet cell depletion in the gastrointestinal tract (Extended Data Fig. [12g](#)), no defect in germ cell maturation and no testicular atrophy (Extended Data Fig. [12h,i](#)) in AU-15330-treated mice—all of which have been reported as toxicities of therapies targeted towards BET proteins^{[28](#)}.

Discussion

We report AU-15330 as a novel, highly specific and VHL-dependent PROTAC degrader of SWI/SNF ATPase components (SMARCA2, SMARCA4 and PBRM1) that shows preferential cytotoxicity in enhancer-binding transcription factor-addicted cancers at low nanomolar concentrations. Our study identifies the SWI/SNF complex as a transcriptional dependency in AR/FOXA1-driven prostate cancer.

Mechanistically, we show that complete inactivation of SWI/SNF ATPase induces a rapid, near-complete and targeted loss of chromatin accessibility at the core-enhancer circuitry of AR, FOXA1, MYC and ERG, thereby attenuating their cancer-promoting transcriptional programs and tempering the enhancer-wired supra-physiologic expression of driver oncogenes (Fig. [4f](#)). These findings are in line with those from recent studies that have used chemical and/or genetic approaches to show that continuous SWI/SNF-remodelling activity is needed to retain enhancers in an open, nucleosome-free conformation^{[29,30](#)}. To our knowledge, this is the first study to

demonstrate that physical chromatin accessibility can be modulated at non-coding regulatory elements as a novel therapeutic strategy in cancer treatment. Thus, recently developed SWI/SNF ATPase inhibitors and degraders add to the growing arsenal of chromatin-targeted therapeutics for directly combating enhancer addiction in human cancers, warranting assessments of their safety and efficacy in clinical trials.

Methods

Cell lines, antibodies, and compounds

Most cell lines were originally obtained from ATCC, DSMZ, ECACC, or internal stock. C4-2B cells were provided by E. Keller (University of Michigan). CWR-R1 cells and a series of enzalutamide-resistant prostate cancer cell lines (LNCaP_Parental, LNCaP_EnzR, CWR-R1_Parental, CWR-R1_EnzR, VCaP_Parental and VCaP_EnzR) were provided by D. Vander Griend (University of Illinois at Chicago)³¹. Bin-67 was generously provided by B. Vanderhyden (Ottawa Hospital Research Institute). All cells were genotyped to confirm their identity at the University of Michigan Sequencing Core and tested routinely for Mycoplasma contamination. LNCaP, 22RV-1, CWR-R1, PC-3, and DU145 were grown in Gibco RPMI-1640 + 10% FBS (ThermoFisher). VCaP was grown in Gibco DMEM + 10% FBS (ThermoFisher). BIN-67 cell lines were grown in custom media (20% fetal bovine serum (FBS), 40% Dulbecco's modified Eagle's medium, 40% Dulbecco's modified Eagle's medium/Ham's F12). Sources of all antibodies are described in Supplementary Table 2. AU-15330 was synthesized by Aurigene (see [Supplementary Text](#)), dBRD9 and VZ 185 were purchased from Tocris Bioscience, and enzalutamide was purchased from Selleck Chemicals.

Computational modelling of AU-15330 - SMARCA2-BD binding

The binding model of AU-15330 in complex with SMARCA2-BD and VHL was generated using Aurigene's proprietary computing algorithm ALMOND (algorithm for modeling neosubstrate degraders). The algorithm is developed using the ICM-Pro integrated modelling platform (http://www.molsoft.com/icm_pro.html) and trained to predict models of ternary complexes of bi-functional molecules with very short or no linkers. The process employs protein-protein docking simulation, exhaustive conformational sampling, small molecule-protein docking, and site-directed scoring of predicted ternary complex models. The computed score estimates the force of induced interactions in the predicted target-E3 ligase complex and is used as a basis for prioritization of degrader binding models. The images were prepared using PyMOL (<https://www.schrodinger.com/products/pymol>).

Cell viability assay

Cells were plated onto 96-well plates in their respective culture medium and incubated at 37 °C in an atmosphere of 5% CO₂. After overnight incubation, a serial dilution of compounds was prepared and added to the plate. The cells were further incubated for 5 days, and the CellTiter-Glo assay (Promega) was then performed according to the manufacturer's instruction to determine cell proliferation. The luminescence signal from each well was acquired using the Infinite M1000 Pro plate reader (Tecan), and the data were analyzed using GraphPad Prism software (GraphPad Software).

Incucyte proliferation assays/Caspase-3/7 green apoptosis assay

A total of 4,000 cells per well were seeded in clear 96-well plates. After overnight incubation, compounds were added to the cells at logarithmic dose series. One day and 8 days after seeding, cellular ATP content was measured using CellTiterGlo (Promega). Measurements after 8 days were divided by the measurement after 1 day (that is, the T0 plate) to derive fold proliferation. For online analysis of cell growth, 4,000 cells per well were seeded in clear 96-well plates (Costar no. 3513). IncuCyte Caspase-3/7 Green Apoptosis Assay Reagent (1:1,000, Essen BioSciences no. 4440) was added, and cells were incubated at 37 °C and 5% CO₂ overnight. On the next day, compounds were added at the desired concentration using the HP digital dispenser D300, and plates were read in an Incucyte ZOOM. Every 2h, phase object confluence (percentage area) for proliferation and green object count for apoptosis were measured. Values for apoptosis were normalized for the total number of cells.

Western blot and immunoprecipitation

Cell lysates were prepared in RIPA buffers (ThermoFisher Scientific) supplemented with cOmpleteTM protease inhibitor cocktail tablets (Sigma-Aldrich), and total protein was measured by Pierce BCA Protein Assay Kit (ThermoFisher Scientific). An equal amount of protein was resolved in NuPAGE 3 to 8%, Tris-Acetate Protein Gel (ThermoFisher Scientific) or NuPAGE 4 to 12%, Bis-Tris Protein Gel (ThermoFisher Scientific) and blotted with primary antibodies. Following incubation with HRP-conjugated secondary antibodies, membranes were imaged on an Odyssey CLx Imager (LiCOR Biosciences). Immunoprecipitations were performed in LNCaP and VCaP cells treated as described. 600 µg of nuclear extracts isolated using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (ThermoFisher Scientific) were immunoprecipitated with SMARCC1, AR, FOXA1, or ERG antibodies according to the manufacturer's protocol. Eluted proteins were subjected to western blot or mass spectrometry analysis. For all immunoblots, uncropped and unprocessed images are provided in Supplementary Figure 1.

RNA isolation and quantitative real-time PCR

Total RNA was isolated from cells using the Direct-zol kit (Zymo), and cDNA was synthesized from 1,000 ng total RNA using Maxima First Strand cDNA Synthesis Kit for PCR with reverse transcription (RT-PCR) (Thermo Fisher Scientific). Quantitative real-time PCR (qPCR) was performed in triplicate using standard SYBR green reagents and protocols on a QuantStudio 5 Real-Time PCR system (Applied Biosystems). The target mRNA expression was quantified using the $\Delta\Delta Ct$ method and normalized to *ACTB* expression. All primers were designed using Primer 3 (<http://frodo.wi.mit.edu/primer3/>) and synthesized by Integrated DNA Technologies. Primer sequences are listed in Supplementary Table 2.

CRISPR knock-out and inducible shRNA knockdown

Guide RNAs (sgRNAs) targeting the exons of human *SMARCA2/BRM* or *SMARCA4/BRG1* were designed using Benchling (<https://www.benchling.com/>). Non-targeting sgRNA, *SMARCA2/BRM* or *SMARCA4/BRG1*-targeting sgRNAs were cloned into lentiCRISPR v2 plasmid according to published literature³²; lentiCRISPR v2 plasmid was a gift from F. Zhang (Addgene plasmid #52961). LNCaP cells were transiently transfected with lentiCRISPR v2 encoding non-targeting or pool of three independent *SMARCA2/BRM* or *SMARCA4/BRG1*-targeting sgRNAs. Twenty-four hours after transfection, cells were selected with 1 μ g ml⁻¹ puromycin for three days. Western blot was performed to examine the knock-out efficiency. The sgRNA sequences are listed in Supplementary Table 2.

ATAC-seq and analysis

ATAC-seq was performed as previously described³³. In brief, 50,000 cancer cells treated with AU-15330 or ZBC-260[30] were washed in cold PBS and resuspended in cytoplasmic lysis buffer (CER-I from the NE-PER kit, Invitrogen, cat. no. 78833). This single-cell suspension was incubated on ice for 5–8 min (depending on the cell line) with gentle mixing by pipetting every 2 min. The lysate was centrifuged at 1,300g for 5 min at 4 °C. Nuclei were resuspended in 50 μ l of 1× TD buffer, then incubated with 2–2.5 μ l Tn5 enzyme for 30 min at 37 °C (Nextera DNA Library Preparation Kit; cat. no. FC-121-1031). Samples were immediately purified by Qiagen minElute column and PCR-amplified with the NEBNext High-Fidelity 2X PCR Master Mix (NEB; cat. no. M0541L) following the original protocol³³. qPCR was used to determine the optimal PCR cycles to prevent over-amplification. The amplified library was further purified by Qiagen minElute column and SPRI beads (Beckman Coulter; cat. no. A63881). ATAC-seq libraries were sequenced on the Illumina HiSeq 2500 (125-nucleotide read length, paired end).

Paired-end .fastq files were trimmed and uniquely aligned to the GRCh38/hg38 human genome assembly using Novoalign (Novocraft) (with the parameters -r None -k -q 13 -k -t 60 -o sam -a CTGTCTCTTATACACATCT), and converted to .bam files using SAMtools (version 1.3.1). Reads mapped to mitochondrial or duplicated reads were removed by SAMtools and PICARD MarkDuplicates (version 2.9.0), respectively. Filtered .bam files from replicates were merged for downstream analysis. MACS2 (2.1.1.20160309) was used to call ATAC-seq peaks. The coverage tracks were generated using the program bam2wig (<http://search.cpan.org/dist/Bio-ToolBox/>) with the following parameters: –pe–rpm–span–bw. Bigwig files were then visualized using the IGV (Broad Institute) open-source genome browser, and the final figures were assembled using Adobe Illustrator.

De novo and known motif enrichment analysis

All de novo and known motif enrichment analyses were performed using the HOMER (v.4.10) suite of algorithms⁴³. Peaks were called by the findPeaks function (-style factor -o auto) at 0.1% false discovery rate; de novo motif discovery and enrichment analysis of known motifs were performed with findMotifsGenome.pl (–size given–mask). The top 10 motifs from the results are shown, and motifs were generally ascribed to the protein family instead of specific family members (unless known).

RNA-seq and analysis

RNA-seq libraries were prepared using 200–1,000 ng of total RNA. PolyA+ RNA isolation, cDNA synthesis, end-repair, A-base addition, and ligation of the Illumina indexed adapters were performed according to the TruSeq RNA protocol (Illumina). Libraries were size selected for 250–300 bp cDNA fragments on a 3% Nusieve 3:1 (Lonza) gel, recovered using QIAEX II reagents (QIAGEN), and PCR amplified using Phusion DNA polymerase (New England Biolabs). Library quality was measured on an Agilent 2100 Bioanalyzer for product size and concentration. Paired-end libraries were sequenced with the Illumina HiSeq 2500, (2 × 100 nucleotide read length) with sequence coverage to 15–20M paired reads.

Libraries passing quality control were trimmed of sequencing adaptors and aligned to the human reference genome, GRCh38. Samples were demultiplexed into paired-end reads using Illumina’s bcl2fastq conversion software v2.20. The reference genome was indexed using bowtie2-build, and reads were aligned onto the GRCh38/hg38 human reference genome using TopHat2³⁴ with strand-specificity and allowing only for the best match for each read. The aligned file was used to calculate strand-specific read count for each gene using HTSeq-count (version 0.13.5)³⁵. EdgeR (version 3.34.1)³⁶ was used to compute differential gene expression using raw read-counts as input. Heatmaps were generated using the ComplexHeatmap³⁷ package in R. For gene

enrichment analysis (GSEA), we first defined ERG and FOXA1 gene signatures from VCaP or LNCaP cells treated with control siRNA or siRNA targeting *ERG*³⁸ or *FOXA1* (generated in this study) containing 250 significantly downregulated genes. For AR and MYC, the Hallmark gene signatures were used. These gene signatures were used to perform a fast pre-ranked GSEA using fgsea bioconductor package³⁹ in R. We used the function fgsea to estimate the net enrichment score and p-value of each pathway, and the plotEnrichment function was used to plot enrichment for the pathways of interest.

ChIP-seq and data analysis

Chromatin immunoprecipitation experiments were carried out using the HighCell# ChIP-Protein G kit (Diagenode) as per the manufacturer's protocol. Chromatin from 5×10^6 cells was used for each ChIP reaction with 10 µg of the target protein antibody. In brief, cells were trypsinized and washed twice with 1× PBS, followed by cross-linking for 8 min in 1% formaldehyde solution. Crosslinking was terminated by the addition of 1/10 volume 1.25 M glycine for 5 min at room temperature followed by cell lysis and sonication (Bioruptor, Diagenode), resulting in an average chromatin fragment size of 200 bp. Fragmented chromatin was then used for immunoprecipitation using various antibodies, with overnight incubation at 4 °C. ChIP DNA was de-crosslinked and purified using the iPure Kit V2 (Diagenode) using the standard protocol. Purified DNA was then prepared for sequencing as per the manufacturer's instructions (Illumina). ChIP samples (1–10 ng) were converted to blunt-ended fragments using T4 DNA polymerase, *Escherichia coli* DNA polymerase I large fragment (Klenow polymerase), and T4 polynucleotide kinase (New England BioLabs (NEB)). A single adenine base was added to fragment ends by Klenow fragment (3' to 5' exo minus; NEB), followed by ligation of Illumina adaptors (Quick ligase, NEB). The adaptor-ligated DNA fragments were enriched by PCR using the Illumina Barcode primers and Phusion DNA polymerase (NEB). PCR products were size-selected using 3% NuSieve agarose gels (Lonza) followed by gel extraction using QIAEX II reagents (Qiagen). Libraries were quantified and quality checked using the Bioanalyzer 2100 (Agilent) and sequenced on the Illumina HiSeq 2500 Sequencer (125-nucleotide read length).

Paired-end, 125 bp reads were trimmed and aligned to the human reference genome (GRC h38/hg38) with the Burrows-Wheeler Aligner (BWA; version 0.7.17-r1198-dirty)⁴⁰. The SAM file obtained after alignment was converted into BAM format using SAMTools (version 1.9). MACS2 (version 2.1.1.20160309) callpeak was used for performing peak calling with the following option: 'macs2 callpeak--call-summits--verbose 3 -g hs -f BAM -n OUT--qvalue 0.05'. For H3K27ac data, the broad option was used. Using deepTools (version 3.3.1) bamCoverage, a coverage file (bigWig format) for each sample was created. The coverage was calculated as the number of

reads per bin, where bins are short consecutive counting windows. While creating the coverage file, the data was normalized with respect to each library size. ChIP peak profile plots and read-density heat maps were generated using deepTools, and cistrome overlap analyses were carried out using the ChIPpeakAnno (version 3.0.0) or ChIPseeker (version 1.29.1) packages in R (version 3.6.0).

HiChIP library preparation and data analysis

HiChIP assay was performed on 5×10^6 DMSO or AU-15330 treated VCaP cells. Frozen cells were resuspended in $1 \times$ PBS and crosslinked with 3 mM DSG and 1% formaldehyde. Washed cells were digested with $0.5 \mu\text{l}$ MNase in $100 \mu\text{l}$ of nuclease digest buffer with MgCl_2 . Cells were lysed with $1 \times$ RIPA, and clarified lysate (approximately 1,400 ng) was used for ChIP. The antibody amount used per ChIP and vendor information are as follows: CTCF: 1.14 μg of Cell Signaling cat. no. 3418; H3K4me3: 3.4 μg of Cell Signaling cat. no. 9751; H3K27ac: 0.4 μg of Cell Signaling cat. no. 8173. The Protein A/G bead pulldown, proximity ligation, and libraries were prepared as described in the Dovetail protocol (Dovetail HiChIP MNase Kit). Libraries were sequenced on an Illumina HiSeq 4000.

Raw fastq files were aligned using BWA mem (version 0.7.17-r1198-dirty) with the -5SP options with an index containing only the main chromosome from the human genome release hg38 (available from the UCSC genome). The aligned paired reads were annotated with pairtools (version 0.3.0) parse (<https://github.com/open2c/pairtools>) with the following options—min-mapq 40—walks-policy 5unique—max-inter-align-gap 30 and the—chroms-path file corresponding to the size of the chromosome used for the alignment index. The paired reads were further processed to remove duplicated reads, sorted with unaligned reads removed with the pairtools sort and the pairtools dedup tools with the basic option to produce an alignment file in the bam format as well as the location of the valid pair. The valid pairs were finally converted to the .cool and .mcool format using the cooler cload and cooler zoomify tools (version 0.8.11)⁴¹ and to the .hic format using the juicer tool (version 1.22.01)⁴².

For the generation of the aggregate peak analyses (APA) plots, we used the HiCExplorer tools (version 3.7) and the hicAggregateContacts command with—range 50000:100000—numberOfBins 30. Plots for all chromosomes were individually computed and summated to generate the global APA plots. The ComplexHeatmaps package³⁷ in R was used for the generation of the final heatmap. For the Hi-ChIP contact heatmap, .hic files were uploaded to the WashU Epigenome Browser (<https://epigenomegateway.wustl.edu/>), and screenshots from gene loci of interest were downloaded using the default viewing conditions.

Super-enhancer analysis

Super-enhancer regions were identified with findPeaks function from HOMER (version v.4.10)⁴³ using options “-style super -o auto”. In addition, the option “-superSlope -1000” was added to include all potential peaks, which were used to generate the super-enhancer plot (super-enhancer score versus ranked peaks). The slope value of greater than or equal to 1 was used to identify super-enhancer clusters. The input files to findPeaks were tag directories generated from alignment files in SAM format with makeTagDirectory function from HOMER.

AU-15330 and enzalutamide formula for *in vivo* studies

AU-15330 was added in 40% of 2-hydroxypropyl- β -cyclodextrin (HP β CD) and sonicated until completely dissolved, and then the solution was further mixed with 5% dextrose in water (D5W) to reach a final concentration of 10% HP β CD. AU-15330 was freshly prepared right before administration to mice. AU-15330 was delivered to mice by intravenous injection either through the tail vein or retro-orbital injection unless otherwise indicated. Enzalutamide was added in 1% carboxymethyl cellulose (CMC) with 0.25% Tween-80 and sonicated until homogenized. Enzalutamide was delivered to mice by oral gavage.

Human prostate tumour xenograft models

Six-week-old male CB17 severe combined immunodeficiency (SCID) mice were procured from the University of Michigan breeding colony. Subcutaneous tumours were established at both sides of the dorsal flank of mice. Tumours were measured at least biweekly using digital calipers following the formula $(\pi/6)(L \times W^2)$, where L is length and W is width of the tumour. At the end of the studies, mice were killed and tumours extracted and weighed. The University of Michigan Institutional Animal Care and Use Committee (IACUC) approved all *in vivo* studies.

For the VCaP non-castrated tumour model, 3×10^6 VCaP cells were injected subcutaneously into the dorsal flank on both sides of the mice in a serum-free medium with 50% Matrigel (BD Biosciences). Once tumours reached a palpable stage (~ 200 mm 3), mice were randomized and treated with either 10, 30 mg kg $^{-1}$ AU-15330, or vehicle through intravenous injection 5 days per week for 3 weeks.

For the VCaP castration-resistant tumour model, 3×10^6 VCaP cells were injected subcutaneously into the dorsal flank on both sides of the mice in a serum-free medium with 50% Matrigel (BD Biosciences). Once tumours reached a palpable stage (~ 200 mm 3), tumour-bearing mice were castrated. Once tumours grew back to the pre-

castration size, mice were randomized and treated with either 60 mg kg⁻¹ AU-15330 or vehicle by intravenous injection 3 days per week, and with or without 10 mg kg⁻¹ enzalutamide by oral gavage 5 days per week for 5 weeks.

For the C4-2B non-castrated tumour model, 1 × 10⁶ cells were injected subcutaneously into the dorsal flank on both sides of the mice in a serum-free medium with 50% Matrigel (BD Biosciences). Once tumours reached a palpable stage (~100 mm³), mice were randomized and treated with either 60 mg kg⁻¹ AU-15330 or vehicle by intravenous injection 3 days per week, and with or without 30 mg kg⁻¹ enzalutamide by oral gavage 5 days per week for 4 weeks. Following the IACUC guidelines, in all treatment arms the maximal tumour size did not exceed the 2.0 cm limit in any dimension and animals with xenografts reaching that size were duly euthanized. The raw tumour volumes and/or weights from all animal efficacy studies are included in the Source Data files.

Prostate patient-derived xenograft models

The University of Texas M. D. Anderson Cancer Center PDX series has been previously described⁴⁴. PDXs were derived from men with CRPC undergoing cystoprostatectomy using described protocols. MDA-PCa-146-12 was derived from a CRPC patient diagnosed with Gleason 5+4=9 prostate adenocarcinoma. MDA-PCa-146-12 was derived from a specimen obtained from the left bladder wall and demonstrated conventional adenocarcinoma (AR⁺). PDXs were maintained in male SCID mice by surgically implanting 2 mm³ tumours coated with 100% Matrigel to both flanks of mice. Once tumours reached ~200 mm³ in size, mice were randomized and divided into different treatment groups receiving either 60 mg kg⁻¹ AU-15330 or vehicle by subcutaneous injection 3 days per week, and with or without 10 mg/kg enzalutamide by oral gavage 5 days per week for 3 weeks. For castration-resistant MDA-PCa-146-12, tumours were established on castrated male SCID mice. Once tumours reached ~100 mm³, mice were randomized and divided into different treatment groups receiving either 60 mg kg⁻¹ AU-15330 or vehicle by intravenous injection 3 days per week, and with or without 30 mg kg⁻¹ enzalutamide by oral gavage 5 days per week for 6 weeks. Following the IACUC guidelines, in all treatment arms the maximal tumour size did not exceed the 2.0 cm limit in any dimension and animals with xenografts reaching that size were duly euthanized. The raw tumour volumes and/or weights from all animal efficacy studies are included in the Source Data files.

Histopathological analysis of organs harvested for drug toxicity

For the present study, organs (liver, spleen, kidney, colon, small intestine, prostate, and testis) were harvested and fixed in 10% neutral buffered formalin followed by embedding in paraffin to make tissue blocks. These blocks were sectioned at 4 µm and stained with Harris haematoxylin and alcoholic eosin-Y stain (both reagents from Leica Surgipath) and staining was performed on Leica autostainer-XL (automatic) platform. The stained sections were evaluated by two different pathologists using a brightfield microscope in a blinded fashion between the control and treatment groups for general tissue morphology and coherence of architecture. A detailed comprehensive analysis of the changes noted at the cellular and sub-cellular level were performed as described below for each specific tissue.

Evaluation of liver

Liver tissue sections were evaluated for normal architecture, and regional analysis for all three zones was performed for inflammation, necrosis, and fibrosis.

Evaluation of spleen

Splenic tissue sections were evaluated for the organization of hematogenous red and lymphoid white pulp regions including necrosis and fibrotic changes if any.

Evaluation of kidney

Kidney tissue sections were examined for changes noted if any in all the four renal functional components, namely glomeruli, interstitium, tubules, and vessels.

Evaluation of colon

Colonic tissue sections were examined for mucosal (epithelium and lamina propria), sub-mucosal, and seromuscular layer changes including crypt changes, goblet cells, inflammatory infiltrate granulation tissue, and mucosal ulceration. A detailed goblet cell evaluation was also performed utilizing Alcian blue staining wherein goblet cells and epithelial cells were counted in ten colonic crypt epithelia in each experimental animal of the various subgroups. Summation of all the goblet and epithelial cells was done, and a ratio of goblet cell to epithelial cell (GC ratio) was calculated per sample.

Evaluation of small intestine

Small intestine tissue sections were examined for mucosal changes such as villous blunting, villous: crypt ratio, and evaluated for inflammatory changes including

intraepithelial lymphocytes, extent (mucosal, sub-mucosal, serosal), and type of inflammatory infiltrate including tissue modulatory effect.

Evaluation of prostate

Prostate tissue sections were evaluated to note for any epithelial abnormality and stromal changes identified in all four lobes (dorsal, anterior, lateral, and ventral). Additionally, any overt inflammatory infiltrate was also examined.

Evaluation of testis

Testicular tissues were examined for the architectural assessment of seminiferous tubules (orderly maturation of germinal epithelial cells devoid of maturation arrest and Sertoli cell prominence), Leydig cells, and interstitial reaction. For an in-depth comprehensive analysis to comment upon the spermatogenesis in a semi-quantitative method, a testicular biopsy score count (Johnsen score) in 100 orderly cross-sections of seminiferous tubules in each animal of all the subgroups at 20 \times magnification was performed. Each of the 100 seminiferous tubules assessed was given a score (score range: 0–10), and the average score was calculated (total sum of score/100).

Alcian blue staining

Alcian blue staining was performed as per the manufacturer's protocol (Alcian Blue Stain Kit (pH 2.5) cat. no. ab150662). Following an overnight incubation of tissue sections at 58 °C, slides were deparaffinized in xylene followed by hydration in ethanol (100%, 70%) and water for 5 min each. Slides were then incubated in acetic acid solution for 3 min followed by a 30 min incubation at room temperature in Alcian blue stain (pH 2.5). Excess Alcian blue was removed by rinsing slides in acetic acid solution for 1 min, and three water washes for 2 min each. Nuclear Fast Red solution was used as a counterstain for 5 min. Slides were subsequently washed in running tap water, dehydrated in ethanol, xylene, and mounted using EcoMount (Thermo Fisher, cat. no. EM897L).

Immunohistochemistry

Immunohistochemistry was performed on formalin-fixed paraffin-embedded 4 μ m sections of mouse or xenograft tissues. Slides with tissue sections were incubated at 58 °C overnight and the next day were deparaffinized in xylene, followed by serial hydration steps in ethanol (100%, 70%) and water for 5 min each. Endogenous tissue peroxidase activity was blocked by placing slides in 3% H₂O₂-methanol solution for 1 h at room temperature. Antigen retrieval was performed by microwaving slides in a

solution of citrate buffer (pH 6) for 15 min, followed by blocking in 2.5% normal horse serum (Vector Laboratories, cat. no. S-2012-50) for 2 h. The slides were then incubated in the following primary antibodies overnight at 4 °C: BRG1 (Abcam cat. no. 108318, 1:100), AR (Millipore cat. no. 06-680, 1:2,000), BRM1 (Millipore Sigma cat. no. HPA029981, 1:100), FOXA1 (Thermo Fisher Scientific cat. no. PA5-27157, 1:1,000), ERG (Cell Signaling Technology cat. no. 97249S, 1:500). ImmPRESS-HRP conjugated anti-mouse–anti-rabbit cocktail from Vector Laboratories (cat. no. MP-7500-50) was used as secondary antibodies (room temperature, 1 h). Visualization of staining was done per the manufacturer’s protocol (Vector Laboratories, cat. no. SK-4100). Following DAB staining, slides were dehydrated in ethanol, xylene (5 min each), and mounted using EcoMount (Thermo Fisher, cat. no. EM897L).

TMT mass spectrometry

VCaP cells were seeded at 5×10^6 cells on a 100 mm plate 24 h before treatment. Cells were treated in triplicate by the addition of test compounds. After 4 h, the cells were harvested and processed by using EasyPep Mini MS Sample Prep Kit (Thermo Fisher, A40006). Samples were quantified using a micro BCA protein assay kit (Thermo Fisher Scientific) and cell lysates were proteolyzed and labelled with TMT 10-plex Isobaric Label Reagent (Thermo Fisher Scientific, 90110) essentially following the manufacturer’s protocol. Briefly, upon reduction and alkylation of cysteines, the proteins were precipitated by adding 6 volumes of ice-cold acetone followed by overnight incubation at 20 °C. The precipitate was spun down, and the pellet was allowed to air dry. The pellet was resuspended in 0.1M TEAB and digested overnight with trypsin (1:50 enzyme:protein) at 37 °C with constant mixing using a thermomixer. The TMT 10-plex reagents were dissolved in 41 ml of anhydrous acetonitrile, and labelling was performed by transferring the entire digest to the TMT reagent vial and incubating it at room temperature for 1 h. The reaction was quenched by adding 8 ml of 5% hydroxylamine and a further 15 min incubation. Labelled samples were mixed together and dried using a vacufuge. An offline fractionation of the combined sample (200 mg) into 10 fractions was performed using high pH reversed-phase peptide fractionation kit according to the manufacturer’s protocol (Pierce, 84868). Fractions were dried and reconstituted in 12 ml of 0.1% formic acid/2% acetonitrile in preparation for LC–MS/MS analysis.

To obtain superior accuracy in quantitation, we employed multinotch-MS3⁴⁵ which minimizes the reporter ion ratio distortion resulting from fragmentation of co-isolated peptides during MS analysis. Orbitrap Fusion (Thermo Fisher Scientific) and RSLC Ultimate 3000 nano-UPLC (Dionex) was used to acquire the data. The sample (2 ml) was resolved on a PepMap RSLC C18 column (75 mm i.d. × 50 cm; Thermo Scientific) at the flowrate of 300 nl min⁻¹ using 0.1% formic acid/acetonitrile gradient system (2–22% acetonitrile in 150 min; 22–32% acetonitrile in 40 min; 20 min wash at

90% followed by 50 min re-equilibration) and direct spray into the mass spectrometer using EasySpray source (Thermo Fisher Scientific). The mass spectrometer was set to collect one MS1 scan (Orbitrap; 60K resolution; AGC target 2×10^5 ; max IT 100 ms) followed by data-dependent, “Top Speed” (3 s) MS2 scans (collision-induced dissociation; ion trap; NCD 35; AGC 5×10^3 ; max IT 100 ms). For multinoth-MS3, top 10 precursors from each MS2 were fragmented by HCD followed by Orbitrap analysis (NCE 55; 60K resolution; AGC 5×10^4 ; max IT 120 ms, 100-500 *m/z* scan range). Proteome Discoverer (v2.1; Thermo Fisher) was used for data analysis. MS2 spectra were searched against SwissProt human protein database (release 11 November 2015; 42,084 sequences) using the following search parameters: MS1 and MS2 tolerances were set to 10 ppm and 0.6 Da, respectively; carbamidomethylation of cysteines (57.02146 Da) and TMT labelling of lysine and N-termini of peptides (229.16293 Da) were considered static modifications; oxidation of methionine (15.9949 Da) and deamidation of asparagine and glutamine (0.98401 Da) were considered variable. Identified proteins and peptides were filtered to retain only those that passed FDR threshold. Quantitation was performed using high-quality MS3 spectra (Average signal-to-noise ratio of 20 and <30% isolation interference).

Meta-analyses of protein interactomes

Interactome proteomics data of AR and ERG was downloaded from published literature^{38,46}. The FOXA1 nuclear co-immunoprecipitation/mass spectrometry experiment was performed in this study as described above. The protein interactomes of AR, ERG, and FOXA1 were ranked based on FDR at the top 10%, and the intersection was taken from these three independent studies.

Assessment of drug synergism

To determine the presence of synergy between two drug treatments, cells were treated with increasing concentrations of either drug for 120 h, followed by the determination of viable cells using the CellTiter-Glo Luminescent Cell Viability Assay (Promega). The experiment was carried out in four biological replicates. The data were expressed as percentage inhibition relative to baseline, and the presence of synergy was determined by the Bliss method using the synergy finder R package⁴⁷.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

All raw next-generation sequencing, ATAC, ChIP, RNA, and HiChIP-seq data generated in this study have been deposited in the Gene Expression Omnibus (GEO) repository at NCBI under accession code [GSE171592](#). [Source data](#) are provided with this paper.

References

1. 1.
Kassabov, S. R., Zhang, B., Persinger, J. & Bartholomew, B. SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol. Cell* **11**, 391–403 (2003).
2. 2.
Kadoch, C. et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
3. 3.
Shain, A. H. & Pollack, J. R. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS ONE* **8**, e55119 (2013).
4. 4.
Bednar, J. et al. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. USA* **95**, 14173–14178 (1998).
5. 5.
Adelman, K. & Lis, J. T. How does Pol II overcome the nucleosome barrier? *Mol. Cell* **9**, 451–452 (2002).
6. 6.
Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
7. 7.
ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

8. 8.

Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

9. 9.

Blackwood, E. M. & Kadonaga, J. T. Going the distance: a current view of enhancer action. *Science* **281**, 60–63 (1998).

10. 10.

Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).

11. 11.

Chen, H. et al. A pan-cancer analysis of enhancer expression in nearly 9000 patient samples. *Cell* **173**, 386–399.e12 (2018).

12. 12.

Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).

13. 13.

Dobersch, S., Rubio, K. & Barreto, G. Pioneer factors and architectural proteins mediating embryonic expression signatures in cancer. *Trends Mol. Med.* **25**, 287–302 (2019).

14. 14.

Northcott, P. A. et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).

15. 15.

Parolia, A. et al. Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. *Nature* **571**, 413–418 (2019).

16. 16.

Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).

17. 17.

Helming, K. C., Wang, X. & Roberts, C. W. M. Vulnerabilities of mutant SWI/SNF complexes in cancer. *Cancer Cell* **26**, 309–317 (2014).

18. 18.

Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* **1**, e1500447 (2015).

19. 19.

Centore, R. C., Sandoval, G. J., Soares, L. M. M., Kadoch, C. & Chan, H. M. Mammalian SWI/SNF chromatin remodeling complexes: emerging mechanisms and therapeutic strategies. *Trends Genet.* **36**, 936–950 (2020).

20. 20.

Farnaby, W. et al. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat. Chem. Biol.* **15**, 672–680 (2019).

21. 21.

Zoppi, V. et al. Iterative design and optimization of initially inactive proteolysis targeting chimeras (PROTACs) identify VZ185 as a potent, fast, and selective von Hippel–Lindau (VHL) based dual degrader probe of BRD9 and BRD7. *J. Med. Chem.* **62**, 699–726 (2019).

22. 22.

Remillard, D. et al. Degradation of the BAF complex factor BRD9 by heterobifunctional ligands. *Angew. Chem. Int. Ed. Engl.* **56**, 5738–5743 (2017).

23. 23.

Mashtalir, N. et al. Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell* **175**, 1272–1288.e20 (2018).

24. 24.

Smith, B. E. et al. Differential PROTAC substrate specificity dictated by orientation of recruited E3 ligase. *Nat. Commun.* **10**, 131 (2019).

25. 25.

Papillon, J. P. N. et al. Discovery of orally active inhibitors of brahma homolog (BRM)/SMARCA2 ATPase activity for the treatment of brahma related gene 1 (BRG1)/SMARCA4-mutant cancers. *J. Med. Chem.* **61**, 10155–10172 (2018).

26. 26.

Viswanathan, S. R. et al. Structural alterations driving castration-resistant prostate cancer revealed by linked-read genome sequencing. *Cell* **174**, 433–447.e19 (2018).

27. 27.

Takeda, D. Y. et al. A somatically acquired enhancer of the androgen receptor is a noncoding driver in advanced prostate cancer. *Cell* **174**, 422–432.e13 (2018).

28. 28.

Faivre, E. J. et al. Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. *Nature* **578**, 306–310 (2020).

29. 29.

Schick, S. et al. Acute BAF perturbation causes immediate changes in chromatin accessibility. *Nat. Genet.* **53**, 269–278 (2021).

30. 30.

Iurlaro, M. et al. Mammalian SWI/SNF continuously restores local accessibility to chromatin. *Nat. Genet.* **53**, 279–287 (2021).

31. 31.

Kregel, S. et al. Acquired resistance to the second-generation androgen receptor antagonist enzalutamide in castration-resistant prostate cancer. *Oncotarget* **7**, 26259–26274 (2016).

32. 32.

Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

33. 33.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

34. 34.

Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12**, R72 (2011).

35. 35.

Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

36. 36.

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

37. 37.

Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

38. 38.

Sandoval, G. J. et al. Binding of TMPRSS2–ERG to BAF chromatin remodeling complexes mediates prostate oncogenesis. *Mol. Cell* **71**, 554–566.e7 (2018).

39. 39.

Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. Preprint at <https://www.biorxiv.org/content/10.1101/060012v3> (2021).

40. 40.

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

41. 41.

Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).

42. 42.

Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).

43. 43.

Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

44. 44.

Palanisamy, N. et al. The MD Anderson prostate cancer patient-derived xenograft series (MDA PCa PDX) captures the molecular landscape of prostate cancer and facilitates marker-driven therapy development. *Clin. Cancer Res.* **26**, 4933–4946 (2020).

45. 45.

McAlister, G. C. et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).

46. 46.

Stelloo, S. et al. Endogenous androgen receptor proteomic profiling reveals genomic subcomplex involved in prostate tumorigenesis. *Oncogene* **37**, 313–322 (2018).

47. 47.

Ianevski, A., He, L., Aittokallio, T. & Tang, J. SynergyFinder: a web application for analyzing drug combination dose-response matrix data. *Bioinformatics* **33**, 2413–2415 (2017).

Acknowledgements

We thank M. Cieslik, E. Young, Y. Cheng and C. Wang from the Michigan Center for Translational Pathology at the University of Michigan, V. Basrur and the Rogel Cancer Center Proteomics Shared Resource, S. Chelur and A. Kumar from Aurigene Discovery Technologies, and T. Dickinson and E. Schulak from Dovetail Genomics for insightful discussions, sharing experimental protocols, and providing technical assistance; S. Wang for providing the BRD4 degrader ZBC260; S. Ellison for editing and proofreading the manuscript; and J. Athanikar for helping with the journal submission. This work was supported by the following mechanisms: Prostate Cancer Foundation (PCF), Prostate Specialized Programs of Research Excellence Grant P50-CA186786, National Cancer Institute Outstanding Investigator Award R35-CA231996, the Early Detection Research Network U01-CA214170, National Cancer Institute P30-CA046592, and the 2020 Movember Distinguished Gentleman's Ride PCF Challenge Award. L.X. is supported by a Department of Defense Prostate Cancer Research Program Idea Development Award (W81XWH-21-1-0500). A.P. is supported by the NIH/NCI F99/K00 pre-doctoral to post-doctoral transition fellowship and the PCF Young Investigator Award. A.M.C. is a Howard Hughes Medical Institute Investigator, A. Alfred Taubman Scholar, and American Cancer Society Professor.

Author information

Author notes

1. These authors contributed equally: Lanbo Xiao, Abhijit Parolia, Yuanyuan Qiao

Affiliations

1. Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI, USA

Lanbo Xiao, Abhijit Parolia, Yuanyuan Qiao, Pushpinder Bawa, Sanjana Eyunni, Rahul Mannan, Sandra E. Carson, Yu Chang, Xiaoju Wang, Yuping Zhang, Josh N. Vo, Steven Kregel, Stephanie A. Simko, Andrew D. Delekta, Mustapha Jaber, Heng Zheng, Ingrid J. Apel, Lisa McMurry, Fengyun Su, Rui Wang, Sylvia Zelenka-Wang, Xuhong Cao, Alexey I. Nesvizhskii, Rohit Mehra & Arul M. Chinnaiyan

2. Department of Pathology, University of Michigan, Ann Arbor, MI, USA

Lanbo Xiao, Abhijit Parolia, Yuanyuan Qiao, Pushpinder Bawa, Sanjana Eyunni, Rahul Mannan, Sandra E. Carson, Yu Chang, Xiaoju Wang, Yuping

Zhang, Josh N. Vo, Steven Kregel, Stephanie A. Simko, Andrew D. Delekta, Heng Zheng, Ingrid J. Apel, Lisa McMurry, Fengyun Su, Rui Wang, Sylvia Zelenka-Wang, Xuhong Cao, Alexey I. Nesvizhskii, Rohit Mehra & Arul M. Chinnaiyan

3. Molecular and Cellular Pathology Program, University of Michigan, Ann Arbor, MI, USA

Abhijit Parolia & Sanjana Eyunni

4. Rogel Cancer Center, University of Michigan, Ann Arbor, MI, USA

Yuanyuan Qiao, Xiaoju Wang, Alexey I. Nesvizhskii, Rohit Mehra & Arul M. Chinnaiyan

5. Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

Josh N. Vo, Alexey I. Nesvizhskii & Arul M. Chinnaiyan

6. Aurigene Discovery Technologies, Electronic City Phase II, Bangalore, India

Sanjita Sasmal, Leena Khare, Subhendu Mukherjee, Chandrasekhar Abbineni, Kiran Aithal, Susanta Samajdar & Murali Ramachandra

7. Dovetail Genomics, Scotts Valley, CA, USA

Mital S. Bhakta, Jay Ghurye & Marco Blanchette

8. Howard Hughes Medical Institute, University of Michigan, Ann Arbor, MI, USA

Xuhong Cao & Arul M. Chinnaiyan

9. Department of Genitourinary Medical Oncology and the David H. Koch Center for Applied Research of Genitourinary Cancers, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Nora M. Navone

10. Department of Internal Medicine/Oncology, University of Michigan, Ann Arbor, MI, USA

Ulka Vaishampayan

11. Vancouver Prostate Centre, Vancouver, British Columbia, Canada

Yuzhuo Wang

12. Department of Urologic Sciences, University of British Columbia, Vancouver, British Columbia, V6T 1Z3, Canada

Yuzhuo Wang

13. Department of Urology, University of Michigan, Ann Arbor, MI, USA

Arul M. Chinnaiyan

Contributions

L.X., A.P., Y.Q. and A.M.C. conceived and designed the studies; L.X. and A.P. performed all of the in vitro and functional genomics experiments with assistance from S.E., S.E.C., H.Z., X.W., S.K., I.J.A. and M.J.; Y.Q. performed all of the animal efficacy studies with assistance from S.A.S. and A.D.D.; A.P. and P.B. carried out all of the bioinformatics analyses with assistance from Y.Z. and J.N.V.; R. Mannan and R. Mehra carried out all of the histopathological evaluations of drug toxicity and quantified all of the histology-based data; S.E. and S.Z.-W. carried out all of the immunohistochemistry with L.M. helping with tissue processing and cross-sectioning. M.S.B., J.G. and M.B. helped with the HiChIP-seq experiment and data analyses. Y.C. helped with modelling drug–protein interaction. F.S. and R.W. generated next-generation sequencing libraries, and X.C. performed the sequencing. S. Sasmal, L.K., S.M., C.A., S. Samajdar, K.A. and M.R. were involved in the discovery, synthesis and initial profiling of the AU-15330 compound. N.M.N., U.V. and Y.W. provided various key preclinical and clinical resources. A.I.N. guided all of the proteomics analyses. L.X., A.P. and A.M.C. wrote the manuscript and organized the final figures.

Corresponding author

Correspondence to [Arul M. Chinnaiyan](#).

Ethics declarations

Competing interests

S. Sasmal., L.K., S.M., C.A., S. Samajdar, K.A. and M.R. are affiliated with Aurigene Discovery Technologies, which is a clinical-stage biotech company with working sites in Bangalore, India and Kuala Lumpur, Malaysia. J.G., M.S.B. and M.B. are affiliated

with Dovetail Genomics, which is an early-stage Santa Cruz-based start-up company developing cutting-edge genomics technologies. A.M.C. is a co-founder and serves on the scientific advisory boards of LynxDx, Oncopia and Esanik. A.M.C. serves on the scientific advisory board of Tempus and Ascentage. The other authors declare no competing interests.

Additional information

Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

Extended data figures and tables

[Extended Data Fig. 1 Conformational model of AU-15330 target interaction and activity profile in diverse cell lines.](#)

(a) Docking model of AU-15330 (cyan sticks) with the SMARCA2 and VHL complex. AU-15330 is suggested to fit into the pocket of SMARCA2 and VHL and capture several key interactions. Key hydrogen bond interactions with protein residues (pink sticks in SMARCA2, white sticks in VHL) are shown by yellow dashes. (b) Effects of AU-15330 (1 μ M, 4h) on the proteome of VCaP cells. Data plotted Log2 of the fold change (FC) versus DMSO (dimethyl sulfoxide) control against –Log10 of the *p-value* per protein (FDR, false discovery rate) from n = 3 independent experiments. All t-tests performed were two-tailed t-tests assuming equal variances. TMT, tandem mass tag. (c) Heatmap showing TMT-based MS abundance of detectable SWI/SNF components after 4h of treatment with AU-15330 at 1 μ M. Data from three independent replicates are shown. (d) Heatmap of relative abundance of several bromodomain-containing proteins detected via Tandem Mass Tag (TMT)-based quantitative MS upon 4h AU-15330 treatment. DMSO, dimethyl sulfoxide (vehicle). (e) Heatmap of mammalian SWI/SNF (BAF) complex subunits split into three constituent modules detected in SMARCC1 (also known as BAF155) nuclear co-immunoprecipitation followed by MS. Direct AU-15330 targets are in bold. (f) Dose-response curves of cells treated with AU-15330 and AU-16235 (inactive epimer of AU-15330). Data are presented as mean +/- SD (n = 6) from one-of-three independent experiments. (g) Crystal violet staining showing the effect of AU-15330 on colony formation. This experiment was repeated independently twice. (h) Dose-response curves and IC₅₀ of cells treated with AU-15330, ACBI1, and BRM014. Data are presented as mean +/- SD (n = 6) from one-of-three independent experiments. (i)

Immunoblots of noted proteins in VCaP cells treated with AU-15330, ACBI1, or BRM014 at increasing concentrations for 24h. Vinculin is the loading control probed on all immunoblots. This experiment was repeated independently twice. (j) Representative immunohistochemistry images showing expression of indicated proteins in patient-derived breast cancer cell lines. (k) Immunoblots of noted proteins in WA-72-P or WA-72-As breast cancer cells treated with DMSO or AU-15330 at noted concentrations for 24h, Vinculin is the loading control probed on a representative immunoblot. This experiment was repeated independently twice.

[Source data](#)

Extended Data Fig. 2 Verification of PROTAC design of AU-15330 and confirmation of on-target growth effects.

(a) Immunoblots for indicated proteins in normal (RWPE) or PCa cells (LNCaP, VCaP, 22RV1, and LAPC4) treated with AU-15330 at varied concentrations. Vinculin is the loading control probed on a representative immunoblot. This experiment was repeated independently twice. (b) Western blot analysis showing the time-dependent effect of AU-15330 on SMARCA2, SMARCA4, and PBRM1 in RWPE, LNCaP, and VCaP cells. Vinculin is the loading control probed on a representative immunoblot. This experiment was repeated independently twice. (c) Immunoblots in LNCaP and VCaP cells examining time-dependent cleavage of PARP upon AU-15330 treatment. Vinculin is the loading control probed on a representative immunoblot. This experiment was repeated independently twice. (d) Dose-response curves of VCaP, LNCaP, PNT2, PNT2, BPH1, Bin67, and HEK293 cells treated with AU-15330, AU-15139, or AU-16235. Data are presented as mean +/- SD (n = 6) from one-of-three independent experiments. (e) Growth curves of non-neoplastic or PCa cells upon treatment with increasing concentrations of AU-15330. Bottom, rightmost panel shows real-time assessment of apoptotic signals in LNCaP cells after treatment with DMSO or increasing AU-15330 concentrations. Data are presented as mean +/- SD (n = 5) from one-of-three independent experiments. (f) (top) Chemical structure of AU-15330, AU-16235 (an epimer control of AU-15330), and AU-15139 (parent bromodomain-binding ligand of AU-15330). (bottom) Immunoblots for SMARCA4 and PBRM1 in LNCaP and VCaP cells treated with AU-15330, AU-15139, or AU-16235 at indicated concentrations. Vinculin is the loading control probed on all immunoblots. This experiment was repeated independently twice. (g) Immunoblots of SMARCA4 and PBRM1 in VCaP and LNCaP cells pre-treated with VL285, MLN4924, bortezomib, or thalidomide for 1h, then treated with AU-15330 at noted concentrations for 4h. Vinculin is the loading control probed on all immunoblots. This experiment was repeated independently twice. (h) Real-time measure showing the rescue effect of VHL ligand on AU-15330-mediated growth inhibition in VCaP and

LNCaP cells. Data are presented as mean \pm SD ($n = 4$) from one-of-three independent experiments.

[Source data](#)

Extended Data Fig. 3 SWI/SNF ATPases, SMARCA2 and SMARCA4, mediate chromatin accessibility at numerous sites across the genome in PCa cells.

(a, b) ATAC-seq read-density heatmaps from VCaP cells treated with DMSO (solvent control), AU-15330, or ZBC-260 (a BRD4 degrader) for indicated durations at genomic sites that are compacted **(a)** or remain unaltered **(b)** upon AU-15330 treatment. Immunoblots show loss of target proteins upon treatment of cancer cells with AU-15330 (1 μ M) for increasing durations or ZBC-260 (10 nM) for 4h. Vinculin is the loading control probed on all immunoblots. This experiment was repeated independently twice. Barplot shows the changes in mRNA expression (RNA-seq) of AU-15330 (1 μ M) target genes in VCaP cells treated for noted durations. **(c)** Schematic outlining the CRISPR/Cas9 and shRNA-based generation of LNCaP cells with either independent or simultaneous inactivation of SWI/SNF ATPases, SMARCA2 and SMARCA4. Immunoblots showing the decrease in target expression in the genetic models shown above. Vinculin is the loading control probed on a representative immunoblot. This experiment was repeated independently twice. **(d)** ATAC-seq read-density heatmaps from genetically engineered LNCaP cells with SMARCA2 and/or SMARCA4 functional inactivation at AU-15330-compacted genomic sites. **(e)** Binding analysis for the regulation of transcription (BART) prediction of specific transcription factors mediating the observed transcriptional changes upon AU-15330 treatment in LNCaP or VCaP cells. The top 10 significant and strong (z-score) mediators of transcriptional responses are labeled (BART, Wilcoxon rank-sum test). **(f)** Top ten de novo motifs (ranked by *p-value*) enriched within AU-15330-compacted genomic sites (HOMER, hypergeometric test) in VCaP cells. **(g)** De novo motif analysis with top 10 motifs (ranked by *p-value*) enriched within genomic sites that retain chromatin accessibility upon AU-15330 treatment in VCaP cells (HOMER hypergeometric test).

[Source data](#)

Extended Data Fig. 4 SWI/SNF inhibition condenses chromatin at enhancer sites bound by oncogenic transcription factors AR and FOXA1 in PCa cells.

(a) ATAC-seq read-density heatmaps from LNCaP cells treated with DMSO or AU-15330 for indicated durations at all genomic sites that lose physical accessibility upon

AU-15330 treatment. **(b)** Genome-wide changes in chromatin accessibility upon AU-15330 treatment for 12h in LNCaP cells, along with genomic annotation of sites that are lost or retained in the AU-15330-treated cells. **(c)** De novo motif analysis with top 10 motifs (ranked by *p-value*) enriched within AU-15330-compacted or unaltered genomic sites in LNCaP cells (HOMER, hypergeometric test). **(d)** ChIP-seq read-density heatmaps for ERG at the AU-15330-compacted genomic sites in VCaP cells after treatment with DMSO or AU-15330 (1 μ M) for indicated times and stimulation with R1881 (1 nM, 3h). **(e)** Genome-wide changes in AR and FOXA1 ChIP-seq peaks upon AU-15330 treatment (1 μ M, 6h) in VCaP cells stimulated with R1881, a synthetic androgen (1 nM, 3h). **(f)** Immunoblots showing the changes in indicated chemical histone marks upon treatment with AU-15330. Vinculin is the loading control probed on a representative immunoblot. This experiment was repeated independently twice. **(g)** ChIP-seq read-density heatmaps for AR, FOXA1, and H3K27Ac at the compacted genomic sites in LNCaP cells after indicated durations of treatment with AU-15330 (1 μ M). **(h)** Genome-wide changes in AR and FOXA1 ChIP-seq peaks upon AU-15330 treatment (1 μ M, 6h) in LNCaP cells stimulated with R1881 (1 nM, 3h). **(i)** ChIP-seq tracks for AR, FOXA1, and H3K27Ac within the *KLK2/3* gene locus in R1881-stimulated VCaP and LNCaP cells with or without AU-15330 (AU).

Extended Data Fig. 5 The SWI/SNF complex is a common chromatin cofactor of the central transcriptional machinery in PCa cells.

(a) The overlap between AR, FOXA1, ERG, and SMARCC1 ChIP-seq peaks in VCaP cells. **(b)** Genomic annotation of oncogenic transcription factor and SWI/SNF (SMARCC1) chromatin binding sites. **(c)** The overlap between transcription factor and SWI/SNF complex shared genomic sites (from a) and H3K27Ac ChIP-seq peaks along with the genomic annotations of the shared binding sites. **(d)** Left: volcano plot showing the AR interacting proteins identified from AR immunoprecipitation followed by MS. Significantly enriched SWI/SNF subunits are highlighted in red (two-sided t-test). Right: Overlap between AR, FOXA1, and ERG interacting proteins identified from in-house or publicly available datasets. **(e)** Immunoblots for indicated proteins followed by nuclear co-immunoprecipitation (IP) of AR, FOXA1, ERG, or SMARCC1 (a core SWI/SNF subunit) in VCaP and LNCaP cells after DHT (dihydrotestosterone) stimulation (10 nM, 3h). This experiment was repeated independently twice.

[Source data](#)

Extended Data Fig. 6 The canonical SWI/SNF complex is the primary cofactor of enhancer-binding transcription factors and is essential for enabling their oncogenic gene programs.

(a, c) Genome-wide ChIP-seq read-density heatmaps and Venn diagrams for CTCF in VCaP (**a**) or LNCaP (**c**) cells treated with either DMSO or AU-15330 (1 μ M) for 6h. Vinculin is the loading control probed on a representative immunoblot. (**b, d**) Immunoblots of indicated proteins in VCaP (**b**) or LNCaP (**d**) cells treated with AU-15330 (1 μ M) for increasing time durations. Total histone H3 is the loading control probed on all immunoblots. This experiment was repeated independently twice. (**e**) GSEA plots for ERG, FOXA1, and MYC-regulated genes using the fold change rank-ordered gene signature from AU-15330-treated (1 nM, 24h) VCaP cells. NES, net enrichment score; adj P, adjusted p-value; DEGs, differentially expressed genes. (**f, g**) GSEA of FOXA1, MYC, or ARID1A-regulated genes (see **Methods** for gene sets) in the fold change rank-ordered AU-15330 gene signature in indicated PCa cells. DEGs, differentially expressed genes. ($n = 2$ biological replicates, GSEA enrichment test) (**h, i**) Expression of indicated genes (qPCR) in VCaP (**h**) or LNCaP (**i**) cells upon treatment with DMSO, AU-15330, dB RD7 (BRD7 degrader), or dB RD7/9 (dual BRD7 and BRD9 degrader) at 1 μ M for 24h. Data are presented as mean \pm SD ($n = 3$, technical replicates) from one-of-two independent experiments. (**j**) Immunoblots for indicated proteins in LNCaP and VCaP cells treated with AU-15330, dB RD9 (BRD9 degrader), or VZ185 (BRD7/9 degrader) at indicated concentrations. Vinculin is the loading control probed on all immunoblots. This experiment was repeated independently twice.

Source data

Extended Data Fig. 7 SWI/SNF inhibition down-regulates the expression of oncogenic drivers through disruption of promoter and super-enhancer interactions.

(a, b) RNA expression (RNA-seq) heatmaps from VCaP or LNCaP cells treated with DMSO, AU-15330 (1 μ M), or ZBC-260 (BRD4 degrader) for the noted durations. $n = 2$, biological replicates. (**c**) RNA expression (qPCR) of indicated genes in stable CRISPR-engineered LNCaP-sgNC (control) or LNCaP-sgSMARCA2 (SMARCA2 inactivation) cells that were treated with a non-target control shRNA or two distinct shRNAs targeting the *SMARCA4* gene. Data are presented as mean \pm SD ($n = 3$, technical replicates) from one-of-two independent experiments. Right, immunoblots showing expression of the indicated protein in CRISPR/shRNA-engineered LNCaP cells. Vinculin is the loading control probed on a representative immunoblot. This experiment was repeated independently twice. (**d**) Normalized read density of AR, FOXA1 and H3K27Ac ChIP-seq signal at the super-enhancer sites ($n = 1,551$ sites) in VCaP cells treated with DMSO or AU-15330 (1 μ M) for 4h or H3K27Ac with 24h AU-15330 (two-sided t-test). For all box plots, the center shows median, box marks quartiles 1–3, and whiskers span quartiles $1–3 \pm 1.5 \times IQR$.

[Source data](#)

[Extended Data Fig. 8 Enhancer-promoter interactions at loci of oncogenic transcription factors with AU-15330.](#)

(a) ATAC-seq and ChIP-seq tracks for AR, FOXA1, and H3K27Ac within the *AR* gene locus in VCaP cells with or without AU-15330 treatment (1 μ M for 6h for AR and FOXA1; 1 μ M for 24h for H3K27Ac). **(b)** H3K27Ac HiChIP-seq heatmaps within the *FOXA1* gene locus in VCaP cells plus/minus treatment with AU-15330 (1 μ M) for 4h (bin size = 25Kb). ATAC-seq read-density tracks from the same treatment conditions are overlaid. Grey highlights mark enhancers, while the blue highlight marks the *FOXA1* promoter. **(c)** Aggregate peak analysis (APA) plots for H3K4me3 (active promoter mark) HiChIP-seq data for all possible interactions between putative enhancers and gene promoters in VCaP cells plus/minus treatment with AU-15330 (1 μ M) for noted durations. **(d)** APA plots for CTCF HiChIP-seq data for all possible interactions between CTCF-bound insulator elements in VCaP cells plus/minus treatment with AU-15330 (1 μ M, 4h). TAD, topologically associating domain. **(e)** CTCF HiChIP-seq heatmaps in a gene locus at Chr14, including the *FOXA1* topologically associating domain (TAD), in VCaP cells plus/minus treatment with AU-15330 (1 μ M) for 4h (bin size = 100Kb). CTCF ChIP-seq read-density tracks from VCaP cells plus/minus AU-15330 treatment (1 μ M) for 6h are overlaid.

[Source data](#)

[Extended Data Fig. 9 AU-15330 is well tolerated in mice and induces on-target degradation of SMARCA2, SMARCA4, and PBRM1.](#)

(a) Immunoblots of indicated proteins in B16F10 and MC38 cells treated with DMSO or AU-15330 (100 nM or 1 μ M). Vinculin is the loading control probed on a representative immunoblot. This experiment was repeated independently twice. **(b)** Schematic outlining the AU-15330 *in vivo* study in non-tumor bearing CD-1 mice. Male mice were treated with vehicle (control) or AU-15330 at the indicated concentration throughout the experiment. **(c)** Pharmacokinetics profile of AU-15330 following intravenous (IV) injection in CD-1 mice. Mice received a single injection at indicated concentration of AU-15330, and plasma levels were determined by HPLC. Data are presented as mean +/- SD (n = 6, biological replicates). **(d)** Immunohistochemistry staining of SMARCA4/BRG1 was carried out using lung, small intestine, and prostate sections after necropsy to show on-target efficacy of AU-15330 *in vivo* (n = 2, biological replicates). **(e)** Body weight measurements showing AU-15330 does not affect weight of non-tumor bearing CD-1 mice. Data are presented as mean +/- SD (n = 6, biological replicates). **(f)** Major organ weight measurements (taken after necropsy) showing AU-15330 does not affect their weight in non-tumor

bearing CD-1 mice. Data are presented as mean \pm SD (n = 6, biological replicates). (g) Complete blood count showing AU-15330 does not affect the hematologic system. Non-tumor bearing CD-1 mice were treated with vehicle or AU-15330 at the indicated concentration throughout the treatment period, and whole blood was then collected and processed. WBC, white blood cells; RBC, red blood cells; PLT, platelets. Data are presented as mean \pm SD (n = 6, biological replicates).

[Source data](#)

Extended Data Fig. 10 Combined *in vivo* treatment with AU-15330 and enzalutamide causes tumor regression in PCa xenografts without toxic effects on other organs.

(a) Schematic outlining the AU-15330 *in vivo* efficacy study using the VCaP-CRPC xenograft model. VCaP cells were subcutaneously grafted in immunocompromised mice that were castrated after 2 weeks of tumor growth to induce disease regression. This was eventually followed by tumor re-growth in the androgen-depleted conditions, generating the aggressive, castration-resistant tumors. (b) Individual tumors and weights from vehicle, enzalutamide, AU-15330, and AU-15330+enzalutamide groups from VCaP-CRPC study (two-sided t-test). Data are presented as mean \pm SEM (vehicle: n = 18, enzalutamide: n = 20, AU-15330: n = 18, AU-15330+enzalutamide: n = 16). For all box plots, the center shows median, box marks quartiles 1–3, and whiskers span the range. (c) Immunoblots of direct AU-15330 targets (upper) and oncogenic transcription factors (bottom) from VCaP-CRPC xenografts (n = 4 tumors/arm) after 5 days of *in vivo* treatment. Vinculin is the loading control probed on a representative immunoblot. (d) Representative immunohistochemistry images from the VCaP-CRPC xenograft study (n = 2 tumors/arm) for SMARCA2 and SMARCA4. (e) Box plot of the percent of cells with positive Ki-67 staining. Two-sided t-test shows significant differences between vehicle vs. enzalutamide, AU-15330, or AU-15330+enzalutamide groups. Data are presented as mean \pm SEM (n = 4, biological replicates). For all box plots, the center shows median, box marks quartiles 1–3, and whiskers span the range. (f) Percent body weight measurement showing the effect of vehicle, enzalutamide, AU-15330, and combination of AU-15330 and enzalutamide throughout the treatment period (two-sided t-test). Data are presented as mean \pm SEM (vehicle: n = 9, enzalutamide: n = 10, AU-15330: n = 9, AU-15330+enzalutamide: n = 8). (g) H&E staining was carried out to examine the effect of AU-15330 *in vivo* using colon, spleen, liver, and kidney sections after necropsy. Representative images of H&E staining are shown. (h) Immunohistochemistry staining of SMARCA4/BRG1 was carried out using liver and kidney sections after necropsy to show on-target efficacy of AU-15330 *in vivo*.

[Source data](#)

Extended Data Fig. 11 AU-15330 inhibits CRPC growth and synergizes with the AR antagonist enzalutamide.

(a) Schematic outlining the AU-15330 *in vivo* efficacy study using the C4-2B (CRPC) xenograft model. C4-2B-xenograft bearing male mice were castrated and, upon tumor regrowth, randomized into various treatment arms. (b) Body weight measurements showing the effect of the indicated treatments on animal weight. Tumor-bearing SCID mice were treated with the indicated drug throughout the treatment period, and the body weight was measured at endpoint. Data are presented as mean +/- SEM (n = 10, biological replicates). (c) Individual tumor volumes from different treatment groups with p-values are shown (two-sided t-test). Data are presented as mean +/- SEM (n = 20, biological replicates). (d) Immunoblots of direct AU-15330 targets (SMARCA2, SMARCA4, and PBRM1) in the whole cell lysate from C4-2B xenografts from all treatment arms after 5 days of *in vivo* treatment (n = 4, biological replicates). Vinculin is the loading control probed on a representative immunoblot. (e-g) VCaP, C4-2B, and LNCaP cells were treated with AU-15330 and/or enzalutamide at varied concentrations to determine the effect on cell growth and drug synergism, with assessments using the Bliss Independence method. Red peaks in the 3D-plots denote synergy with the average synergy scores noted above. The mean of three biological replicates is shown on top. Data are presented as mean (n = 4) from one-of-three independent experiments. (h) Crystal violet staining showing the synergistic effect of AU-15330 and enzalutamide on colony formation in VCaP and LNCaP. (i, j) Dose-response curves of VCaP cells treated with enzalutamide in combination with DMSO or AU-15330 at indicated concentrations. Data are presented as mean +/- SD (n = 4) from one-of-three independent experiments. (k) Dose-response curves of VCaP_Parental and VCaP_EnzR cells treated with enzalutamide or AU-15330. Data are presented as mean +/- SD (n = 6) from one-of-three independent experiments. (l) IC₅₀ for AU-15330 in enzalutamide-resistant (EnzR) LNCaP and VCaP cells after 5 days of treatment.

Source data

Extended Data Fig. 12 AU-15330 inhibits tumor growth of an enzalutamide-resistant patient-derived xenograft (PDX) model without evident toxicities.

(a) Schematics outlining the AU-15330 *in vivo* efficacy studies using the MDA-PCa-146-12 (top) or the MDA-PCa-146-12-CRPC (bottom) xenograft model. MDA-PCa-146-12-CRPC xenograft-bearing male mice were castrated and, upon tumor regrowth, randomized into various treatment arms that were administered vehicle, enzalutamide, or the combination of AU-15330+enzalutamide at indicated concentrations. (b) Tumor

volume measurements (caliper twice per week) showing efficacy of enzalutamide alone or in combination with AU-15330 in the enzalutamide-resistant MDA-PCa-146-12 PDX model ($n = 20$ /arm; two-sided t-test). Data are presented as mean \pm SEM (vehicle: $n = 18$, enzalutamide: $n = 18$, AU-15330+enzalutamide: $n = 16$). (c) Individual tumor weights from different treatment groups from the MDA-PCa-146-12 PDX study with p -values indicated (two-sided t-test). Data are presented as mean \pm SEM (vehicle: $n = 18$, enzalutamide: $n = 18$, AU-15330+enzalutamide: $n = 8$). (d) Waterfall plot showing percent change from baseline of individual tumors from the MDA-PCa-146-12-CRPC model with indicated treatment group after 43 days of treatment. (e, f) Animal body weight measurements showing the effect of vehicle, enzalutamide, and combination of AU-15330 and enzalutamide on animal weight in the (e) MDA-PCa-146-12 or the (f) MDA-PCa-146-12-CRPC PDX models. Tumor-bearing SCID mice were treated with vehicle, enzalutamide, or a combination of AU-15330 and enzalutamide at the indicated concentration throughout the treatment period. Data are presented as mean \pm SEM (for e, vehicle: $n = 9$, AU-15330: $n = 9$, AU-15330+enzalutamide: $n = 8$; for f, vehicle: $n = 7$, AU-15330: $n = 8$, AU-15330+enzalutamide: $n = 8$). (g) Representative Alcian blue staining images from the large intestinal tract harvested at the VCaP-CRPC efficacy study endpoint ($n = 2/treatment\ group$). Right, quantification of goblet:epithelial cell densities in the colon (two-sided t-test). Data are presented as mean \pm SEM ($n = 6$, biological replicates). (h) Top, Representative H&E of the testis gland harvested from DMSO or AU-15330-treated intact male mice after 21 days of *in vivo* treatment. Right, quantification of germ cell density and maturation carried out using the Johnsen scoring system (two-sided t-test). Bottom, gross images of the testis glands. Data are presented as mean \pm SEM ($n = 6$, biological replicates). For all box plots, the center shows median, box marks quartiles 1–3, and whiskers span the range. (i) Individual testes weight and images from different treatment groups of the C4-2B xenograft efficacy study at endpoint (i.e., after 24 days of treatment) with p -values indicated (two-sided t-test). Data are presented as mean \pm SEM (vehicle: $n = 9$, enzalutamide: $n = 10$, AU-15330: $n = 10$, AU-15330+enzalutamide: $n = 10$).

[Source data](#)

Supplementary information

[Supplementary Information](#)

This file contains Supplementary Text and Supplementary Figure 1 (uncropped western blots from all figures).

[Reporting Summary](#)

[Peer Review File](#)

[Supplementary Table 1](#)

Antiproliferative half-maximal inhibitory concentrations of AU-15330 across cancer cell lines.

[Supplementary Table 2](#)

Antibodies, qPCR primers, CRISPR–Cas9 single guide RNA sequences, short hairpin RNA sequences, and compounds used in this study.

[Supplementary Table 3](#)

Effects of AU-15330 on hERG potassium current. Raw measurement values from three independent experiments are included. E-4031 is an experimental class III antiarrhythmic drug that blocks hERG-type potassium channels and was used as a positive control. N, number of cells tested; SEM, standard error of the mean; SD, standard deviation; +1 = whitish discoloration (slight) along the flow of the stock solution as it disperses in the external buffer and it becomes clear without any visible particles upon shaking the contents.

Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

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

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

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

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

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

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

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

[**Source Data Extended Data Fig. 9**](#)

[**Source Data Extended Data Fig. 10**](#)

[**Source Data Extended Data Fig. 11**](#)

[**Source Data Extended Data Fig. 12**](#)

Rights and permissions

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

Xiao, L., Parolia, A., Qiao, Y. *et al.* Targeting SWI/SNF ATPases in enhancer-addicted prostate cancer. *Nature* **601**, 434–439 (2022). <https://doi.org/10.1038/s41586-021-04246-z>

- Received: 17 March 2021
- Accepted: 15 November 2021

- Published: 22 December 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04246-z>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04246-z>

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

- Article
- Open Access
- [Published: 18 November 2021](#)

Genome surveillance by HUSH-mediated silencing of intronless mobile elements

- [Marta Seczynska](#)¹,
- [Stuart Bloor](#)¹,
- [Sergio Martinez Cuesta](#) [ORCID: orcid.org/0000-0001-9806-2805](#)² &
- [Paul J. Lehner](#) [ORCID: orcid.org/0000-0001-9383-1054](#)¹

Nature volume 601, pages 440–445 (2022)

- 12k Accesses
- 200 Altmetric
- [Metrics details](#)

Subjects

- [Chromatin](#)
- [Gene silencing](#)
- [Retrovirus](#)
- [Transposition](#)
- [Virus–host interactions](#)

Abstract

All life forms defend their genome against DNA invasion. Eukaryotic cells recognize incoming DNA and limit its transcription through repressive chromatin modifications. The human silencing hub (HUSH) complex transcriptionally represses long interspersed element-1 retrotransposons (L1s) and retroviruses through histone H3 lysine 9 trimethylation (H3K9me3)^{1,2,3}. How HUSH recognizes and initiates silencing

of these invading genetic elements is unknown. Here we show that HUSH is able to recognize and transcriptionally repress a broad range of long, intronless transgenes. Intron insertion into HUSH-repressed transgenes counteracts repression, even in the absence of intron splicing. HUSH binds transcripts from the target locus, prior to and independent of H3K9me3 deposition, and target transcription is essential for both initiation and propagation of HUSH-mediated H3K9me3. Genomic data reveal how HUSH binds and represses a subset of endogenous intronless genes generated through retrotransposition of cellular mRNAs. Thus intronless cDNA—the hallmark of reverse transcription—provides a versatile way to distinguish invading retroelements from host genes and enables HUSH to protect the genome from ‘non-self’ DNA, despite there being no previous exposure to the invading element. Our findings reveal the existence of a transcription-dependent genome-surveillance system and explain how it provides immediate protection against newly acquired elements while avoiding inappropriate repression of host genes.

[Download PDF](#)

Main

The mammalian genome is under constant threat from invasion by mobile genetic elements including transposons and viruses. Controlling this activity is fundamental to genome integrity. These defence strategies often use repressive chromatin to silence target gene expression and major chromatin-silencing factors in mammalian cells include: (1) small RNA guides complementary to nascent transcripts and (2) sequence-specific DNA-binding proteins⁴. PIWI-interacting RNAs (piRNAs) guide PIWI proteins to transposon transcripts and promote repressive chromatin at germline transposon loci⁵. piRNAs are derived from piRNA clusters, genomic loci enriched in transposon-derived sequences^{6,7}. The piRNA pathway therefore relies on the memory of transposon invasions to provide adaptive, sequence-based immunity. The large KRAB-containing zinc-finger protein (KRAB-ZFP) family of sequence-specific DNA-binding proteins recruit TRIM28 and the SETDB1 methyltransferase to deposit H3K9me3 heterochromatin at target loci^{7,8}. piRNA and KRAB-ZFP pathways are mostly active in the germ line and pluripotent stem cells, whereas the HUSH complex silences mobile elements in pluripotent stem cells and differentiated cells. HUSH represses evolutionary young L1 retrotransposons^{2,3}, the only active autonomous mobile transposons in humans, as well as integrated lentiviruses¹ and unintegrated murine retroviral DNA via NP220⁹. The importance of HUSH in controlling lentiviral infection is emphasised by the finding that complex primate lentiviruses encode accessory proteins (Vpr and Vpx) that degrade HUSH^{10,11,12}.

To silence mobile elements, the HUSH complex of TASOR, MPP8 and periphilin, recruits two effectors: MORC2—an ATP-dependent chromatin remodeler—enables chromatin compaction^{13,14}, and SETDB1 deposits H3K9me3¹. The chromodomain of MPP8 binds to H3K9me3-modified chromatin anchoring HUSH at the target locus. However, how HUSH recognizes its targets to initiate H3K9me3 deposition is unknown.

Intronless transgenes are HUSH-repressed

Since HUSH-repressed L1s are found in diverse genomic integration sites^{2,3,15}, the signal for HUSH recognition must be intrinsic to the L1. To confirm that the L1 sequence confers HUSH repression independent of its integration site, we expressed a lentiviral fluorescent reporter encoding the L1 open reading frame (ORF) and a P2A-iRFP cassette. L1 expression was monitored by flow cytometry with iRFP fluorescence reflecting L1 mRNA abundance (Extended Data Fig. 1a). Inactivation of the ORF2 endonuclease¹⁶ (D205A mutation) reduces retrotransposition; the reporter thus monitors expression from initial L1 integrations (Extended Data Fig. 1c,d). Lentiviral L1 reporter ($L1_{lenti}$) expression is repressed within the entire wild-type population (Fig. 1a), and disrupting HUSH by knockout of HUSH subunits or by TASOR degradation by lentiviral Vpx^{10,11,12} restores $L1_{lenti}$ expression, whether the reporter is integrated before or after HUSH disruption (Fig. 1a, Extended Data Fig. 1b, e–g). As $L1_{lenti}$ is expressed from most integration sites following HUSH depletion (Fig. 1a, Extended Data Fig. 1e,g), HUSH-mediated L1 silencing is independent of integration site. Lentiviruses predominantly integrate in transcribed gene bodies¹⁷, whereas the piggyBac transposase directly integrates at randomly distributed TTAA sites¹⁸. L1 reporter expression from an inducible, piggyBac transposon vector ($L1_{pb}$) confirmed HUSH-dependent repression from most integration sites (Fig. 1b, Extended Data Fig. 1h) and HUSH-mediated H3K9me3 deposition that led to decreased RNA Poll II occupancy and reporter mRNA levels (Fig. 1c). The signal for HUSH repression is therefore intrinsic to L1 and independent of the mechanism and site of genome integration.

Fig. 1: Diverse intronless transgenes are HUSH-repressed.

 **figure 1**

a, Repression of L1 reporter lentivirus in wild-type (WT) (black) or *TASOR*-knockout (KO) (purple) HeLa cells, measured by flow cytometry. **b, c**, L1 reporter integrated by piggyBac transposase. **b**, Doxycycline (Dox)-induced expression in wild-type and *TASOR* KO HeLa cells measured by flow cytometry. CMV, cytomegalovirus promoter. **c**, Chromatin immunoprecipitation with quantitative PCR (ChIP–qPCR) assays of H3K9me3 (left; mean of $n = 2$ biological replicates \pm s.d.) and RNA polymerase II (Pol II) (middle; mean of $n = 3$ biological replicates \pm s.d.) in wild-type and *TASOR* KO HeLa cells at the reporter. L1 transcript levels assayed by quantitative PCR with reverse transcription (RT–qPCR) (right; mean of $n = 3$ technical replicates \pm s.d.). **d**, Doxycycline-induced expression of piggyBac reporter without ORF2 sequence (left) and with ORF2 sequence (4 kb) replaced by 4 \times ORF1 (4 \times 1 kb in size) (right) integrated into wild-type or *TASOR* KO HeLa cells. **e**, HUSH-mediated repression of GFP lentiviral reporters bearing different untranslated cDNA sequences measured by

flow cytometry 72 h after transduction. Length of the cDNA sequence is indicated in brackets and fold change of reporter expression in *TASOR* KD and wild-type cells measured by geometric mean fluorescence is indicated on the graph. Frequency is normalized to mode (**a, b, d, e**).

HUSH restriction of L1 retrotransposition depends on the native nucleotide sequence of the L1 ORF². By testing the HUSH sensitivity of reporters bearing single L1 ORFs¹⁹ (ORF1 or ORF2), we found that the ORF2 sequence alone is responsible for HUSH-mediated repression of L1 (Fig. [1d](#), left, Extended Data Fig. [1i](#)). However, replacing the 4-kb ORF2 with 4 tandem repeats of the 1-kb ORF1 also caused HUSH repression (Fig. [1d](#), right, Extended Data Fig. [1j](#)), suggesting that HUSH repression is not unique to ORF2.

We therefore tested the HUSH sensitivity of lentiviral transgenes with different DNA sequences (Fig. [1e](#), Extended Data Fig. [2a–c](#)). To exclude effects on mRNA translation, we inserted DNA sequences lacking an ATG start codon, with a single-nucleotide frameshift at the 3'-untranslated region (3' UTR) of the GFP reporter (Fig. [1e](#)). Diverse, integrated transgenes containing cDNA sequences from a wide range of human genes were all HUSH-repressed (Fig. [1e](#)), as were transgenes entirely ‘foreign’ to the human genome, for example, the bacterial Cas9 nuclease (Fig. [1e](#), Extended Data Fig. [2d](#)). HUSH therefore silences sequence-diverse self and foreign mobile genetic elements, the latter being important as it excludes the possibility of ‘genetic memory’. HUSH-mediated transgene repression was maintained over multiple cell divisions, was independent of the number of transgene integrations and showed a significant correlation with the length of inserted DNA (Extended Data Fig. [2e–g](#)). While the L1 ORF1 (1kb) reporter is HUSH-insensitive, tandem repeats of ORF1 gradually acquire HUSH repression as their size increases (Fig. [1d](#), Extended Data Fig. [3h](#)). Transgene length therefore contributes to HUSH susceptibility, with short (up to 1 kb) transgenes most likely to escape HUSH-mediated repression (for example, L1 ORF1, iRFP or a fragment of Xist long noncoding RNA (lncRNA)) (Fig. [1d](#), Extended Data Fig. [2h](#)). However, lentiviral reporters encoding short 1-kb fragments of ORF2 (or 3-kb ORF2 deletion mutants) remained HUSH-repressed (Extended Data Fig. [2i–k](#)), indicating a role for nucleotide composition in HUSH targeting.

We found no correlation between HUSH-mediated repression and adenine and thymine (AT) sequence content (Extended Data Fig. [3a](#)), and decreasing the overall AT content of ORF2 did not alleviate HUSH-dependent silencing (Extended Data Fig. [3b](#), [e](#)). However, HUSH-mediated repression strongly correlates with the A nucleotide content of the sense strand (Extended Data Fig. [3c](#)), with ORF2 showing a strong A (41%) versus T (20%) bias in the sense strand²⁰. Indeed, a reverse-complement ORF2 reporter is completely HUSH-resistant, despite expressing a full-length transcript (Extended Data Fig. [3d–g](#)). The HUSH complex therefore represses a broad range of

invading DNAs, with transgene length and high A content in the sense strand acting as key determinants of HUSH targeting.

Transcription is required for repression

Chromatinization of invading DNA precedes genome integration²¹. We therefore investigated whether HUSH initiates repression prior to transgene integration. Lentiviral ORF2 transgenes were HUSH-repressed in both the absence and presence of raltegravir, an inhibitor of viral integration (Fig. 2a, left, Extended Data Fig. 4a, b). Furthermore, transfected lentiviral plasmids encoding (1) ORF2, (2) synthetic ORF2 or (3) Cas9 cDNA sequences were HUSH-repressed (Fig. 2a, right, Extended Data Fig. 4c), as were non-viral plasmids (Extended Data Fig. 4d), indicating that HUSH can initiate silencing prior to transgene integration.

Fig. 2: HUSH binds target RNA and initiates silencing before DNA integration.

 **figure 2**

a, HUSH-mediated repression of non-integrated reporters. Left, HUSH-mediated repression of integrated and non-integrated GFP reporter lentiviruses with no insert (empty) or with synthetic ORF2 measured by flow cytometry 24 h after transduction and calculated as the ratio of reporter expression in wild-type and *TASOR* knockdown

(KD). Data are mean of $n = 3$ biological replicates \pm s.d.; two-sided *** $P = 0.002$, ** $P = 0.008$ versus corresponding no-insert sample, unpaired t -test with Welch's correction. Right, flow cytometry histograms showing expression from GFP lentiviral plasmids containing different untranslated sequences transfected into wild-type or *TASOR* KD 293T cells. gMFI, geometric mean fluorescence intensity. **b**, Top, genome browser track depicting input and H3K9me3 chromatin immunoprecipitation with sequencing (ChIP-seq) signal over the unique fragment of the SFFV-driven or promoterless L1 reporter integrated into wild-type and *TASOR* KO HeLa cells. Bottom, ChIP-qPCR quantifying H3K9me3 and total histone H3 levels at a SFFV-driven or promoterless L1 lentiviral reporter integrated into wild-type and *TASOR* KO HeLa cells. Data are mean of $n = 3$ biological replicates (independent polyclonal integrations of the reporters) \pm s.d.; *** $P = 0.0006$, ** $P = 0.002$, * $P = 0.003$ versus wild-type promoter, paired two-tailed t -test. Red arrows indicate position of the primers used for subsequent quantitative PCR. **c**, RIP in *SETDB1* KO 293T cells with haemagglutinin (HA) tag knocked into *TASOR* or *PPHLN1* locus, showing periphilin and TASOR association with the indicated RNAs (see Extended Data Fig. [4k–m](#) for more details). Data are mean \pm s.d.; $n = 3$ independent experiments, normalized to input. **d**, Enrichment of periphilin RIP sequencing (RIP-seq) peaks at different repetitive elements in *SETDB1* KO (mix) cells. *SETDB1* KO (mix) is a polyclonal cell pool after *SETDB1* CRISPR–Cas9. Significant enrichment is defined as a fold change score above 1 with empirical Benjamini–Hochberg adjusted one-sided P -values (q); *** $q = 0.0002$. **e**, Genome browser tracks depicting periphilin and control RIP-seq signal over intronic L1 elements in wild-type and *SETDB1* KO (mix) cells.

Source data

HUSH targets endogenous, full-length, young L1s that are often enriched within transcriptionally permissive euchromatin, suggesting a role for transcription in HUSH targeting^{2,3,15}. To directly test whether transcription is required to initiate HUSH-mediated silencing, we transduced HeLa cells with either the standard, spleen focus forming virus (SFFV) promoter-driven L1_{lenti} reporter or an otherwise identical promoterless reporter. HUSH-dependent H3K9me3 accumulated over the transcriptionally active L1 reporter, but was significantly reduced in the absence of a promoter (Fig. [2b](#), Extended Data Fig. [4e–h](#)). Deletion of the promoter region from *TAF7*, an endogenous HUSH target gene also reduced transcription (Extended Data Fig. [4i](#), right) and locus-specific H3K9me3 deposition (Extended Data Fig. [4i](#), left, j) confirming that transcription is required to both initiate and maintain H3K9me3 over HUSH-sensitive loci. Furthermore, silencing cannot be conferred solely by the DNA sequence, as the sequences of HUSH-sensitive and HUSH-insensitive transgenes are identical.

A transcriptional requirement in HUSH-mediated silencing suggests that HUSH binds reporter RNA. Native RNA immunoprecipitation (RIP) showed that periphilin specifically binds RNA from a HUSH-sensitive reporter but not from a HUSH-resistant reporter (Fig. 2c, Extended Data Fig. 4k, l). Notably, these results in SETDB1-deficient cells indicate that HUSH must bind reporter RNA prior to and independent of H3K9me3 deposition (Extended Data Fig. 4m). Transcription is therefore required for transgene repression, and periphilin binding to transgene RNA is likely to contribute to its recognition by HUSH.

To gain a global view of RNAs bound by endogenous periphilin (Extended Data Fig. 5a, b), we performed UV-cross-linked RIP and genome-wide analysis. Periphilin binding showed a significant overlap with genomic repeats, with specific enrichment over L1 elements (Fig. 2d, e, Extended Data Fig. 5c, d, f). There was no significant overlap between periphilin peaks and other repeat classes, with only transcripts of the Tigger DNA transposon family showing significant binding (Extended Data Fig. 5d, f). Periphilin preferentially bound transcripts from full-length, evolutionary young L1s (Extended Data Fig. 5e), reflecting the selective, genome-wide, HUSH-mediated H3K9me3 deposition over these L1 elements^{2,3}, as well as from other HUSH-targeted loci (Extended Data Fig. 5f, right). Periphilin recognition of nascent RNA therefore specifies target loci for HUSH repression.

Introns protect against HUSH repression

We next investigated why transcribed cDNA sequences, but not their endogenous genomic loci (Fig. 1e, Extended Data Fig. 6a), are HUSH-repressed. A key difference is that coding regions of neither cDNAs nor L1s are separated by long intragenic non-coding DNA regions (that is, introns) prompting us to investigate whether HUSH sensitivity was intron-dependent.

We compared HUSH repression of: (1) an intronless reporter in which iRFP is followed by non-coding ORF2 (iRFP-ORF2) and (2) an otherwise identical reporter with the second intron of human β-globin (*HBB* IVS2) inserted within the iRFP (Fig. 3a). Intron insertion abrogates HUSH-mediated repression (Fig. 3a, Extended Data Fig. 6b), and HUSH-mediated repression was also abolished by intron insertion at the 5' or 3' end of ORF2 (Fig. 3c, Extended Data Fig. 7c, d). Insertion of an antisense GFP ‘stuffer’ sequence had no effect (Fig. 3c, Extended Data Fig. 7a). This loss of HUSH repression was associated with decreased periphilin binding (Extended Data Fig. 7g–h) and decreased H3K9me3 deposition (Fig. 3b, Extended Data Fig. 7b). Intron-mediated HUSH protection was also observed for: (1) ORF2 reporters of different architecture expressed from an inducible, piggyBac transposon vector, (2) Cas9 reporters expressed from the piggyBac transposon vector, and (3) lentiviral reporters (Extended Data Fig. 6c–e), and was lost following Cre–lox-mediated deletion of an

intron sequence from the integrated transgene, implying that the intron is required continuously to maintain protection (Extended Data Fig. [6f](#)).

Fig. 3: Introns protect against HUSH, even in the absence of intron splicing.

 [figure 3](#)

HUSH-mediated repression of intronless and intron-containing iRFP-ORF2 piggyBac reporters. **a, b**, Second intron from the human β -globin gene (*HBB* IVS2) cloned within the iRFP gene. **a**, Flow cytometry histograms showing expression in wild-type and *TASOR* KO HeLa cells. **b**, ChIP–qPCR quantification of H3K9me3 and total histone H3 at reporters in wild-type and *TASOR* KO HeLa cells. Data are mean of $n = 3$ independent experiments \pm s.d.; ** $P < 0.008$, * $P = 0.02$ versus intronless wild type, ratio-paired two-tailed *t*-test. **c**, HUSH-mediated repression of reporter with intron(s) or control sequence cloned at the 5' or 3' of ORF2, measured by flow cytometry and shown as the ratio of reporter expression in *TASOR* KO and wild-type cells. Data are mean from n biological replicates \pm s.d.; *** $P \leq 0.0001$, one-way analysis of variance (ANOVA) post hoc pairwise comparisons versus no-intron condition with Bonferroni correction. **d**, Flow cytometry histograms showing expression from reporters with different introns from human genes or control sequences cloned at the 5' end of ORF2. Intron size is shown in parentheses. BFP, blue fluorescent protein. **e**, Flow cytometry histograms showing expression from reporters with different *HBB* IVS2 mutant introns cloned 5' of ORF2. Gel image (right) confirms that mutant introns are not spliced from the reporter. ss, splice site. **f**, Quantification of HUSH-mediated repression of reporters from Fig. [3d, e](#), Extended Data Fig. [8c](#) by flow cytometry and calculated as the ratio of reporter expression in *TASOR* KO and wild-type HeLa cells. Data are mean of n biological replicates (independent polyclonal integrations of the reporters) \pm s.d.; *** $P \leq 0.0001$, * $P = 0.044$, ** $P = 0.009$, one-way ANOVA post hoc pairwise comparisons versus intronless with Bonferroni correction. asGFP data is the same as in 5' control from **c**.

Four additional human introns (*EEF1A1*, *NXF1*, *SMC5* and *ACTB*) cloned into the iRFP-ORF2 reporter also provided protection from HUSH-mediated repression (Fig. 3d, f, Extended Data Fig. 7e), an effect not seen with a small artificial intron (chimeric β -globin–IgG), or reporters with similar-length control ‘stuffer’ sequences. The reduction in HUSH sensitivity correlated with the length of intron (Fig. 3f, Extended Data Fig. 7f). The *SMC5* intron, despite being poorly spliced, prevented HUSH-mediated repression more effectively than fully spliced *HBB* and *EEF1A1* introns (Fig. 3d, f, Extended Data Fig. 7e), suggesting that intron excision by the splicing machinery may not be required for protection against HUSH repression. To investigate whether splicing is required for intron-mediated protection, we generated a series of *HBB* IVS2 5' and 3' splice-site mutants (Extended Data Fig. 8a) which, despite effectively abolishing splicing (mutants no. 1 and no. 2), counteracted HUSH-mediated reporter repression as effectively as the wild-type intron (Fig. 3e, f). Mutant intron no. 1 has a 5' splice-site deletion critical for early spliceosome assembly at the transcript²², suggesting that intron-mediated HUSH protection is independent of assembly of the core spliceosome at the transgene RNA. *HBB* IVS2 splice mutants with either a 3' splice-site mutation or deletion of the last 60 nucleotides (including the branch-point site that pairs with the 5' splice site to form a splicing intermediate) not only counteracted HUSH, but provided more effective protection from HUSH-mediated repression than the wild type intron (Fig. 3e, f, Extended Data Fig. 8a, b). Therefore, even in the absence of splicing, introns protect transgenes against HUSH-mediated repression, whereas effectively spliced stuffer sequences flanked by a 5' splice site, a branch point and a 3' splice site, did not counteract HUSH (Fig. 3f, Extended Data Fig. 8c). Thus it is the intron itself rather than the splicing process that protects against HUSH-mediated repression.

HUSH targets endogenous intronless loci

Our data suggest that HUSH provides a genome-surveillance system to repress diverse transcribed, intronless invading DNAs, and predict that genomic loci from similar invading DNAs are bound and silenced by HUSH. Such loci include retrogenes and processed pseudogenes, created when reverse-transcribed cellular mRNA integrates into the genome, as part of a retrotransposition event²³. We detected HUSH binding and HUSH-mediated H3K9me3 at the loci of transcribed processed pseudogenes and retrogenes, but not on their intron-containing, transcribed parent genes (Fig. 4a, Extended Data Fig. 9a–c, e). Many HUSH-repressed pseudogenes and retrogenes are positioned within transcriptionally active genes, similar to HUSH-regulated L1s. The *MAB21L2* retrogene—a non-transcribed parologue of the HUSH-repressed *MAB21L1* retrogene—is not HUSH-repressed, confirming the critical requirement for transcription in HUSH-mediated repression (Extended Data Fig. 9d).

Fig. 4: Transcribed processed pseudogenes and retrogenes are bound and silenced by the HUSH complex.

 figure 4

a, Visualization of HUSH-dependent H3K9me3, HUSH–MORC2-occupancy and RNA sequencing in wild-type and *HUSH* KO K562 cells at representative loci of

processed pseudogene and retrogene. Genome browser tracks are generated from publicly available BigWig files³⁵. Arrowheads indicate transcriptional direction of the gene. **b**, Volcano plots showing \log_2 fold change (\log_2 FC) of periphilin over control RIP-seq-normalized read counts for three gene categories: processed pseudogenes (left), intronless (middle) and intron-containing protein-coding genes (right); representative data from *SETDB1* KO (mix) cells. For each data point, significance was determined after a comparative assessment of counts between conditions using negative binomial generalized linear models as implemented in edgeR. Multiple testing correction of significance was performed using the false discovery rate (FDR) method; $n = 4$ independent experiments. Only genes with periphilin RIP-seq signal greater than 0.3 RPKM are included (>0.3 reads per kilobase of transcript, per million mapped reads (RPKM) in each RIP-seq replicate from both *SETDB1* KO (mix) and wild-type 293T cells). Genes with periphilin peaks overlapping L1 elements are excluded. Intronless protein-coding genes are defined as genes that produce only intronless isoforms. ZNFs, zinc finger family genes. **c**, Schematic of genome surveillance by the HUSH complex. HUSH recognizes long, intronless mobile DNA and targets it for transcriptional silencing. Host genes are protected against HUSH by the presence of introns (left of DNA strand). An average human protein-coding gene contains ten 6,355-bp-long introns³⁵. Transcription of the target initiates HUSH-mediated repression: periphilin binds its specific target transcript, MPP8 recruits SETDB1 to deposit H3K9me3. Periphilin–RNA and MPP8–H3K9me3 interactions anchor HUSH at the target locus (area with dashed outline, right).

Similarly, periphilin bound only retrotranscribed and not intron-containing parent genes (Extended Data Fig. 10a–c). Genomic analysis revealed that 20% of transcribed, non-L1-overlapping pseudogenes and 17% of intronless genes showed at least twofold enrichment of the periphilin RIP signal (Fig. 4b, Extended Data Fig. 10d). There was no enrichment of periphilin binding over intron-containing genes (Extended Data Fig. 10d), with the 5% of genes with bound periphilin predominantly containing HUSH-repressed long (over 2 kb) exons or zinc-finger family (ZNF) members as seen for HUSH-dependent H3K9me3^{1,15} (Fig. 4b, Extended Data Figs. 9f, 10e). HUSH repression of processed pseudogenes and retrogenes—all bona fide endogenous mobile elements—emphasises the physiological role of HUSH in defending the genome against invading retroelements.

Discussion

Our study reveals how the HUSH epigenetic repressor complex provides a versatile defence system against genome invasion. Without previous exposure to its targets, HUSH is able to recognize and transcriptionally repress a broad range of sequence-diverse, intronless DNAs, whereas intron-containing DNAs are resistant to HUSH-

mediated repression. The defining feature of HUSH targets is therefore the presence of long, intronless transcription units, an intrinsic feature of retroelements, including L1 retrotransposons. Non-reverse-transcribed, intronless invading DNAs are also targeted for repression, including transfected cDNA plasmids. HUSH is therefore ‘programmed’ to control the spread of integrating, RNA-derived mobile elements within the host genome, representing a universal, cell-autonomous genome-surveillance system (Fig. 4c). The HUSH-mediated repression of endogenous L1s^{2,3,15} is a consequence of this programming rather than a recognition of unique L1 sequences. Genomic evidence for HUSH repression of sequence-diverse, retrotransposition-derived, endogenous genes supports this conclusion and validates our findings with reporter genes. HUSH specificity for target length and A-rich bias in the sense strand may reflect retroviral reliance on ‘structurally poor’ A-rich RNA sequences to support viral cDNA synthesis during reverse transcription²⁴ and may therefore allow a more selective targeting of reverse-transcribed elements. Moreover, HUSH silencing of transgenes, including most cDNAs larger than about 1.5 kb, explains why many cDNAs remain difficult to express, a practical problem in both gene therapy and in ectopic gene expression in cultured cells.

The dependence of HUSH-mediated repression on transcription is reminiscent of transcription-coupled heterochromatin formation in *Schizosaccharomyces pombe*²⁵, where, as with HUSH, transcription is required for both the initiation and propagation of H3K9me3. -The association of periphilin with its target RNAs even in the absence of H3K9me3 deposition provides support for a critical role of RNA in HUSH-mediated repression. Binding of periphilin to nascent RNA provides specificity for target recognition by recruiting and stabilizing HUSH at target loci independent of the MPP8 chromodomain, and enables HUSH to respond to increased transcription if H3K9me3 levels decline, such as during cell division. Similar to *S. pombe*, transcription-induced recruitment of HUSH to replicated chromatin may ensure inheritance of the repressed state following DNA replication^{26,27}. This requirement for active transcription explains preferential targeting of full length L1s in euchromatic environments by HUSH, and conversely, why HUSH ignores older, degenerate L1s that have lost transcriptional activity^{2,3,15}.

Importantly, intron-mediated protection from HUSH-mediated silencing does not require efficient intron splicing or spliceosome recruitment. Given the complex network of RNA-binding proteins involved in exon–intron definition and splicing²⁸, intronic sequences may counteract HUSH by recruiting proteins other than core splicing factors that compete with periphilin for transcript binding. Alternatively, HUSH may be sensitive to nucleosome distribution, with the increased occupancy over exons versus introns^{29,30} correlating with reduced elongation rates^{29,30,31}. Slow elongation through long exons may trigger HUSH recruitment, which is counteracted by the decreased nucleosome density and increased elongation in cellular introns,

consistent with HUSH-mediated H3K9me3 deposition over long exons of endogenous genes. Shorter introns are much less likely to affect nucleosome positioning (with each nucleosomes occupying 147 nt) than longer introns, consistent with the limited or absent HUSH protection afforded by the short *ACTB* and very short artificial intron. The well-recognized ability of introns to enhance gene expression (intron-mediated enhancement) can, at least in part, be explained by the capacity of introns to protect transgenes from HUSH-mediated silencing³².

To distinguish self from non-self, the host immune system recognizes conserved molecular patterns that are maintained in invading pathogens but are absent from the host. Most mammalian genes are organized such that exons comprise small islands within a sea of intronic sequences, whereas the cDNA products of reverse transcription are RNA-derived and intronless. Long, intronless cDNA, the product of reverse transcription, is therefore the molecular pattern recognized by HUSH, which provides a means to distinguish invading retroelements from host genes. Thus, HUSH comprises a component of the innate immune system. To avoid HUSH recognition, retroelements would need to maintain long, non-coding intron sequences, but are constrained by selective pressure for a compact genome. Bypassing the restriction imposed by HUSH therefore poses a major challenge. Whereas retroviral transcripts are often spliced, the intervening sequences are coding sequences and very different from the classical long non-coding introns of cellular genes. Consequently, primate lentiviruses evade HUSH by encoding accessory proteins that degrade HUSH^{10,11,12}, whereas endogenous retroelements are unable to evade HUSH activity.

The innate immune response provides immediate defence but does not confer long-lasting immunity. HUSH-selective targeting of evolutionary young L1s^{2,3} suggests a limited ability to provide long-lasting repression over evolutionary timescales. By contrast, DNA sequence-specific KRAB-ZFPs are less agile in repressing young retroelements, as it takes several million years to evolve a KRAB-ZFP with high affinity for a new DNA sequence^{33,34}. By rapidly repressing transcription of novel retroelements without the need for genetic memory, HUSH buffers any potentially deleterious effects on cellular fitness. This gives the host a time window to establish sequence-specific adaptive repression to effectively restrict these retroelements and may facilitate their domestication.

Methods

Plasmids

A list and details of all plasmids used in the study are in Supplementary Table 1.

Cell culture

HeLa cells were obtained from ECACC and HEK 293T and Jurkat cells were from ATCC. Cell morphology was assessed for authentication. All cell lines were grown in IMDM plus 10% FCS and penicillin/streptomycin (100 U ml⁻¹). Cell cultures were routinely tested and found to be negative for mycoplasma infection (MycoAlert, Lonza).

Antibodies

Antibodies for immunoblotting: rabbit anti-TASOR (Atlas, HPA006735, 1:5,000), rabbit anti-MPP8 (Proteintech, 16796-1-AP, 1:5,000), rabbit anti-periphilin1 (Sigma-Aldrich, HPA038902, 1:5,000), rabbit anti-MORC2 (Bethyl Laboratories, A300-149A, 1:5,000), rabbit anti-SETDB1 (Proteintech, 11231-1-AP; 1:5,000), rat anti-haemagglutinin (HA) tag (3F10, Sigma-Aldrich, 11867423001, 1:10,000), mouse anti-β-actin peroxidase conjugate (Sigma-Aldrich, A3854; 1:20,000), mouse anti-p97 (Abcam, ab11433, 1:5,000), rabbit anti-α-tubulin (11H10, CST, 2125, 1:5,000). Horseradish peroxidase (HRP)-conjugated secondary antibodies for immunoblotting were obtained from Jackson ImmunoResearch: Peroxidase AffiniPure Goat Anti-Mouse IgG (H+L) (115-035-146, 1:10,000), Peroxidase AffiniPure Goat Anti-Rabbit IgG (H+L) (111-035-144, 1:10,000), Peroxidase AffiniPure Goat Anti-Rat IgG (H+L) (112-035-143, 1:10,000). Antibodies for intracellular staining for flow cytometry: mouse anti-HA tag Alexa Fluor 647 conjugate (Cell Signaling, 3444; 1:50; used only for PPHLN1–HA and HA–TASOR knockin validation). Antibodies for ChIP–qPCR: rabbit anti-H3K9me3 (Abcam, ab8898) 5 µg per immunoprecipitation, rabbit anti-histone H3 (Abcam, ab1791) 5 µg per immunoprecipitation and rabbit anti-RNA Pol II (Bethyl Laboratories, A304-405A, 7.5 µg per immunoprecipitation).

CRISPR–Cas9 mediated gene disruption

HeLa or HEK 293T cells were transfected with a pool of sgRNAs cloned into a Cas9-containing plasmid (pSpCas9(BB)-2A-Puro) using TransIT HeLa Monster or TransIT 293T (Mirus) according to the manufacturer's protocol. Transfected cells were enriched with 24 h of puromycin selection (2 µg ml⁻¹) starting 24 h after transfection. HeLa *TASOR* KD, HEK 293T *TASOR* KD and *SETDB1* KO (mix) cell lines were maintained as mixed KO populations. *HUSH*, *SETDB1* and *MORC2* KO HeLa cells were generated as described^{1,13} and are polyclonal KO populations derived from a HeLa clone harbouring a repressed GFP reporter (pHRSIN-pSFFV-GFP-WPRE-P_{GK}-Zeo^R) integrated at pericentromeric site on chromosome 7: 57848728 (hg19). Parental HeLa cells are GFP⁻ and *HUSH*, *SETDB1* and *MORC2* KO cells are GFP⁺ because of de-repression of the GFP reporter.

Lentiviral production and transduction

Lentivirus was produced by transfecting HEK 293T cells with the lentiviral vector plus the packaging plasmids pCMVΔR8.91 and pMD2.G using TransIT-293 transfection reagent (Mirus). The viral supernatant was collected 48 h later, cell debris was removed with a 0.45- μm filter and target cells transduced by spin infection at 1,800 rpm for 60 min. Transduced HeLa cells were selected with the following drug concentrations: puromycin, 2 $\mu\text{g ml}^{-1}$; hygromycin, 100 $\mu\text{g ml}^{-1}$; and blasticidin, 5 $\mu\text{g ml}^{-1}$. For experiments with non-integrated virus, cells were transduced in the presence of 1 μM raltegravir.

For the ‘one-pot’ establishment assay, WT HeLa cells were initially transduced with lentiviral vector encoding mCherry (pHRSIN-pSFFV-mCherry-WPRE) at a multiplicity of infection (MOI) <1 and mCherry⁺ cells were purified by fluorescence-activated cell sorting (FACS), resulting in 98% pure mCherry⁺ populations (Supplementary Figure 2). mCherry⁺ WT and mCherry⁻ *TASOR* KD cells were mixed at a 1:1 ratio and transduced with the lentiviral GFP reporters by spin infection. Reporter expression was typically analysed 2, 4 and 6 days after transduction by flow cytometry. Gating strategy is depicted in Extended Data Fig. 2c, Supplementary Fig. 2. Reciprocal mixing (mCherry⁺ *TASOR* KD and mCherry⁻ WT) was used to validate results.

Transfection

WT mCherry⁺ and *TASOR* KD mCherry⁻ HEK 293T cells were mixed at a 1:1 ratio and transfected using TransIT-293T (Mirus) according to the manufacturer’s protocol.

PiggyBac-mediated integration of reporter constructs

HeLa or HEK 293T cells were co-transfected with pB-transposon plasmid and piggyBac transposase-expression plasmid at 5:1 or 2.5:1 ratio using TransIT-HeLa Monster or TransIT-293T (Mirus). Transfected cells were selected with blasticidin (5 $\mu\text{g ml}^{-1}$) for at least 3 days starting from 2 days after transfection. For flow cytometry assays, two cell lines were mixed at a 1:1 ratio prior to transfection. For assays with GFP reporters, WT mCherry⁺ cells were mixed with *TASOR* KD mCherry⁻ HeLa cells. For assays with iRFP reporters, WT GFP⁻ HeLa cells were mixed with *TASOR* KO GFP⁺ HeLa cells, which both harbour additional a HUSH-sensitive GFP reporter at chr7:57848728 (hg19). See Supplementary Figure 2 for gating strategies in flow cytometry analyses. Reporter expression was typically analysed 7 and 12 days after transfection and was induced by plating cells in media with doxycycline (1 $\mu\text{g ml}^{-1}$) 24 h prior to flow cytometry analysis or ChIP-qPCR.

Flow cytometry

Live cells were analysed on a LSR Fortessa (BD). Data were analysed using FlowJo v10.6.1 (LCC) software. Cell sorting was carried on a FACSaria Fusion (BD).

Immunoblotting

Cells were lysed in 100 mM Tris pH 7.4 with 1% SDS followed by boiling and vortexing to shear genomic DNA. Lysates were then boiled in SDS sample buffer, separated by SDS-PAGE and transferred to PVDF membranes (Millipore). Membranes were probed with the indicated antibodies and reactive bands visualised with ECL, Supersignal West Pico or West Dura (Thermo Scientific).

CRISPR–Cas9 mediated knock-in of HA tag

For C-terminal periphilin tagging, the HA sequence was inserted upstream of the stop codon at the *PPHLN1* endogenous locus via CRISPR homology-directed repair. For N-terminal TASOR tagging, HA was inserted downstream of the *TASOR* start codon. Single-stranded donor oligonucleotides (ssODN) were used as donor templates and purchased from IDT. HEK 293T cells were transfected with single guide RNA (sgRNA) plasmid (pSpCas9(BB)-2A-Puro) and single-stranded donor template. Transfected cells were enriched by puromycin selection and single-cell cloned. Clonal populations were screened for the presence of HA tag by intracellular flow cytometry staining using anti-HA antibody. The genetic modifications were validated by PCR on genomic DNA followed by sequencing. sgRNA and ssODN sequences listed in Supplementary Table 1.

CRISPR–Cas9 mediated deletion of *TAF7* promoter

Prior to the modification of the *TAF7* locus, HeLa cells were transduced with lentivirus encoding codon-optimized C-terminally HA-tagged TAF7 (*TAF7*_(opt)-HA) and blasticidin resistance as a selection marker. *TAF7* is an essential gene³⁶ and stable expression of exogenous *TAF7*_(opt) was used to compensate for the loss of expression from the endogenous *TAF7* locus due to promoter deletion. Sequence was codon-optimized so that exogenous *TAF7*_(opt) was not detected in RT-qPCR or ChIP-PCR.

Two sgRNAs targeting the *TAF7* promoter region were cloned into pSpCas9(BB)-2A-Puro (PX459, V2.0): one targeting within the first 80 nucleotides of the *TAF7* 5' UTR and a second approximately 850 nt upstream of the transcription start site. Two sgRNA plasmids were mixed at a 1:1 ratio and transfected into HeLa *TAF7*_(opt)-HA-

expressing cells. Twenty-four hours later, cells were treated with puromycin ($2 \mu\text{g ml}^{-1}$) for 24 h and single-cell cloned 5 days after transfection. The genetic deletion effects were validated by PCR on genomic DNA and loss of *TAF7* expression measured by RT-qPCR. Sequences of primers and sgRNAs are detailed in Supplementary Table 1.

Chromatin immunoprecipitation

Cells were cross-linked in 1% formaldehyde for 10 min, quenched in 0.125 M glycine for 5 min and lysed in cell lysis buffer (1 mM HEPES, 85 mM KCl and 0.5% NP-40). Nuclei were pelleted by centrifugation and then lysed in nuclear lysis buffer (5 mM Tris, 10 mM EDTA and 1% SDS) for 10 min. The chromatin was sheared with a Bioruptor (Diagenode Pico) to obtain a mean fragment size of <300 bp. Insoluble material was removed by centrifugation. The chromatin solution was diluted to a final SDS concentration of 0.1% and precleared with Pierce Protein G magnetic beads (Thermo Fisher) and then immunoprecipitated overnight with 5 µg primary antibody and Protein G-magnetic beads. Beads were washed twice with low-salt buffer (20 mM Tris pH 8.1, 2 mM EDTA, 50 mM NaCl, 1% Triton X-100, 0.1% SDS), once with high-salt buffer (20 mM Tris pH 8.1, 2 mM EDTA, 500 mM NaCl, 1% TritonX-100, 0.1% SDS), once with LiCl buffer (10 mM Tris pH 8.1, 1 mM EDTA, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate) and twice with TE. Protein-DNA complexes were eluted in 150 mM NaHCO₃ and 1% SDS at 65 °C. Cross-links were reversed by overnight incubation at 65 °C with 0.3 M NaCl and RNase A. Proteinase K was then added, the samples were incubated for 2 h at 45 °C, and then the DNA was purified with a spin column (Qiagen PCR Purification Kit). Quantification by qPCR was performed on a QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific) using SYBR green PCR mastermix (Thermo Fisher Scientific). qPCR primer sequences are detailed in Supplementary Table 1.

For ChIP-seq, immunoprecipitated DNA was subjected to library preparation (NEBNext Ultra II DNA Library Prep Kit, Illumina). Libraries were purified, quantified, multiplexed (with NEBNext Multiplex Oligos for Illumina kit, E7335S) and sequenced with 2× 50-bp pair-end reads on Illumina Novaseq platform (Genomics Core, Cancer Research UK Cambridge Institute).

Bioinformatics data processing and analyses were performed using Bash (v4.2.46), R (v3.6) and Python (v3.8.5) programming languages as well as the following tools: FastQC (Babraham Bioinformatics) (v0.11.7) cutadapt³⁷ (v1.16), HISAT2³⁸ (v2.1.0), SAMtools³⁹ (v1.9), sambamba⁴⁰ (v0.6.6) and deepTools⁴¹ (v3.1.0). Raw fastq files were quality checked with FastQC and trimmed with cutadapt to remove adapter sequences and low-quality base calls (quality score < 20). Depending on the experiment, the resulting reads were aligned using HISAT2 to either the human

reference genome only (version GRCh38) or the human reference genome concatenated with the sequence of the unique fragment from reporter construct (P2A-iRFP), duplicates were marked using sambamba and alignments were formatted using SAMtools. BigWig files containing genomic signal were computed at single -base resolution and normalized to counts per million (CPM) using deepTools. Further details are available in the GitHub page of this study (<https://github.com/semacu/hush>).

Native RIP-qPCR

Reporter expression was induced by doxycycline ($1 \mu\text{g ml}^{-1}$) for 24 h prior to the experiment. Cells were lysed in HLB-N buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM MgCl₂ and 0.5% NP-40), incubated on ice for 5 min and lysate was underlaid with 1/4 volume of HLB + NS (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM MgCl₂, 0.5% NP-40 and 10% (wt/vol) sucrose). Nuclei were pelleted by centrifugation (420g, 5 min) and then lysed in RIP buffer (25 mM Tris pH 7.4, 150 mM KCl, 5 mM EDTA, 0.5 mM DTT, 0.5% NP-40 and 100 U ml⁻¹ SUPERase-IN). The nuclear fraction was sonicated (Diagenode Pico) and insoluble material was removed by centrifugation (8,000g, 10 min). The nuclear fraction was immunoprecipitated with Pierce anti-HA magnetic beads (Thermo Fisher) for 2 h at 4 °C. Beads were washed four times with RIP buffer and RNA was extracted from beads (and input samples) using TRIzol and standard phenol-chloroform extraction. The aqueous phase containing the RNA was loaded onto RNeasy mini columns (QIAGEN) with 2 volumes of 100% ethanol and RNA was purified according to the manufacturer's protocol. RNA was on-column DNase I treated and reverse transcribed using random hexamers and SuperScript III Reverse Transcriptase (Thermo Fisher Scientific). Quantification by qPCR was performed on QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific) using SYBR green PCR mastermix (Thermo Fisher Scientific). qPCR primers sequences are detailed in Supplementary Table 1.

UV-crosslinked RIP-seq

Cells were UV treated (254 nM UV-C at 0.3 J cm⁻²) in PBS, lysed in HLB-N buffer, incubated on ice for 5 min and lysate was then underlaid with 1/4 volume of HLB + NS. Nuclei were pelleted by centrifugation (420xg, 5 min) and lysed in RIP buffer: (25mM Tris pH 7.4, 150mM KCl, 5mM EDTA, 0.5mM DTT, 0.5% NP-40 and 100 U/ml RNasin (Promega)). The nuclear fraction was sonicated (Diagenode Pico), treated with TURBO-DNase (4U), and insoluble material was removed by centrifugation (8,000g, 10 min). The nuclear fraction was immunoprecipitated with Pierce anti-HA magnetic beads (Thermo Fisher) for 2 h at 4 °C. Beads were washed once with RIP buffer, once with RIP buffer + TURBO-DNase (2U), 2× RIPA buffer

(50 mM Tris pH 7.4, 100 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS), 1× high-salt RIPA (50 mM Tris-HCl pH 7.4, 500 mM NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate), 1× low-salt wash (15 mM Tris-HCl pH 7.4, 5 mM EDTA), for 5 min each time at room temperature with rotation. Beads were digested with proteinase K in proteinase K buffer (50 mM Tris-Cl (pH 7.5), 100 mM NaCl, and 1 mM EDTA, 0.25% SDS) and RNA was isolated by standard phenol-chloroform extraction. RNA from the first RIP-seq experiment in *SETDB1* KO (mix) was in addition rigorously treated with TURBO-DNaseI prior to library preparation. Immunoprecipitated RNA was subjected to DNA library preparation using SMARTer Stranded Total RNA-Seq Kit V3—Pico Input Mammalian (Takara Bio) according to the manufacturer’s instructions with initial fragmentation at 94 °C for 3 or 4 min and ribosomal RNA depletion step included. The library quality was determined using Bioanalyzer, and sequenced on Illumina MiniSeq platform as paired-end 32-bp and 43-bp reads using MiniSeq High-Output 75 cycles kit.

Bioinformatics data processing and analyses were performed using Bash, R (v3.6.0) and Python (v3.8.5) programming languages as well as the following tools: FastQC (Babraham Bioinformatics) (v0.11.7), UMI-tools⁴² (v1.1.1), cutadapt³⁷ (v1.16), HISAT2 (v2.1.0)³⁸, SAMtools (v1.9)³⁹, deepTools⁴¹ (v3.1.0), BEDTools⁴³ (v2.30.0), data.table (v1.13.2), GenomicFeatures⁴⁴ (v1.38.2), edgeR^{45,46} (v3.28.1), and GAT⁴⁷ (v1.0). Raw fastq files were quality checked with FastQC, unique molecular identifiers extracted using UMI-tools and resulting reads trimmed with cutadapt. Alignments to the human reference genome (version GRCh38) were performed with HISAT2, then formatted and deduplicated using SAMtools and UMI-tools respectively. Peaks were called using a customised approach involving BEDTools, deepTools, several Bash commands, datatable and edgeR. Genomic repeats were obtained from RepeatMasker and L1Base^{48,49} and associations with the RIP-seq peaks were investigated using GAT and BEDTools. Tables integrating gene information, RIP-seq signal and repeats were obtained using BEDTools, data.table, GenomicsFeatures and edgeR. Finally combined bigWig files containing genomic signal were prepared with SAMTools and computed at single base resolution and normalized to CPM using deepTools. More details available in the GitHub page of this study <http://github.com/semacu/hush>.

Northern blot

Sample preparation, agarose gel separation and transfer to the membrane were all performed using a NorthernMax Kit (Invitrogen) according to the manufacturer’s recommendation. In brief, 1–10 µg of sample RNA or 2 µg Millennium Markers (Invitrogen) were suspended in formaldehyde loading dye and loaded onto a 6-mm-thick 1% Agarose-LE gel and run at 5 V cm⁻¹ (150 V, 110 min) in 1× MOPS running buffer. The samples were transferred to a BrightStar-Plus positively charged nylon membrane (Invitrogen) over 120 min, via the described downward transfer apparatus

stacked on paper towels. Following transfer, the membrane was UV (254 nm) cross-linked using 120 mJ energy (Stratagene, Stratalinker 1800) and photographed under UV to record the marker positions (Invitrogen, iBright CL1000 Imaging System). Following a 30 min, 68 °C, prehybridization in ULTRAhyb ultrasensitive hybridization buffer, the membrane was incubated overnight at 68 °C with 100 pM digoxigenin-labelled RNA probes, directed against iRFP (nucleotides 4–300) and *ACTB* (nucleotides 69–618 of mRNA, NM_001101). Membrane was washed with 1× low stringency wash solution (room temperature) and 2× NorthernMax high stringency wash buffer (68 °C), prior to blocking at room temperature with 1× casein blocking buffer (Sigma-Aldrich). The membrane was incubated for 60 min with 50 mU ml⁻¹ anti-digoxigenin-POD (poly), Fab fragments (Roche) in 1× blocking buffer, followed by 4 washes in 1× PBS + 0.1% Tween 20 and visualised using a SuperSignal West chemiluminescent substrate (Thermo Fisher) and the Invitrogen, iBright CL1000 Imaging System.

Primers used to generate PCR amplicons against the indicated regions of each gene are listed in Supplementary Table 1. The amplicons were used in a T7 polymerase reaction substituting the NTPs for DIG RNA labelling mix (Roche), to generate antisense digoxigenin labelled RNA probes. The reaction was digested with TURBO DNase (Invitrogen) for 15 min at 37 °C, before purification using an RNeasy MinElute cleanup kit (Qiagen).

RT-qPCR

Total RNA was extracted using the RNeasy Plus Mini kit (Qiagen) with on-column DNase I treatment according to the manufacturer's instructions. RNA was reverse transcribed into cDNA using an equimolar mixture of random hexamers and oligo (dT)₁₆ primers by SuperScript III Reverse Transcriptase (Thermo Fisher Scientific). RNA quantification was performed using the $\Delta\Delta C_t$ method and normalized against *ACTB* or *GAPDH* transcript levels. Primer sequences are detailed in Supplementary Table 1.

Analysis of splicing

Efficiency of splicing of the reporter transcripts were determined by semi-quantitative PCR using intron-flanking primers (see Supplementary Table 1) detecting both unspliced and spliced reverse-transcribed mRNA. cDNA was prepared as for RT-PCR. Corresponding plasmids served as DNA controls.

Statistics and reproducibility

Statistical details, including the statistical test used, type (one- or two-sided), adjustments for multiple comparison and sample sizes (n), are reported in the figures and figure legends. The following figure panels show representative data from at least two independent experiments that showed similar results: Fig. 3e, Extended Data Figs. 1b, e, i, 2a, g, k, 3a, c, d, 4d, l, m, 5b, f, 6b, f, 7g. The following figure panels show representative data from at least three independent biological replicates that showed similar results: Figs. 1d, 2a, right, 3a, d, Extended Data Figs. 1f, 2d, h, 4b, c, k, 6c, 7c, e, h, 8b, c. The following figure panels show representative data from at least four independent biological replicates that showed similar results: Fig. 1a, b, e, Fig. 2e, Extended Data Figs. 2d, 3b, f, 10a, b, e. The experiments in Fig. 1d and Extended Data Fig. 1c were performed once, but were internally controlled for both positive and negative results. The Northern blot experiments in Extended Data Figs. 1h, j, 3g, 4f, 7d were performed once, but were internally controlled for both positive and negative results. The ChIP-seq experiments in Fig. 2b (top) and Extended Data Fig. 4h, j were performed once, but the results were independently validated by two independent ChIP-qPCR experiments.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Gels and blots source images are provided in Supplementary Fig. 1. Next-generation sequencing data have been deposited at the Gene Expression Omnibus (GEO) with accession number [GSE181113](#). The publicly available data² are available at GEO under accession number GSE95374 (ChIP-seq and RNA-sequencing data on the HUSH complex). The version of the human reference genome used in this study is GRCh38 (GENCODE v35, <https://www.gencodegenes.org/human/>). Repeats were obtained from RepeatMasker (v UCSC hg38) and L1Base^{48,49}. [Source data](#) are provided with this paper.

Code availability

For details about the bioinformatics data analyses, check the GitHub page for this study at <http://github.com/semacu/hush>.

References

1. 1.

Tchasovnikarova, I. A. et al. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* **348**, 1481–1485 (2015).

2. 2.

Liu, N. et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**, 228–232 (2018).

3. 3.

Robbez-Masson, L. et al. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res.* **28**, 836–845 (2018).

4. 4.

Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).

5. 5.

Aravin, A. et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207 (2006).

6. 6.

Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744–747 (2007).

7. 7.

Ninova, M., Tóth, K. F. & Aravin, A. A. The control of gene expression and cell identity by H3K9 trimethylation. *Development* **146**, dev181180 (2019).

8. 8.

Ecco, G., Imbeault, M. & Trono, D. KRAB zinc finger proteins. *Development* **144**, 2719–2729 (2017).

9. 9.

Zhu, Y., Wang, G. Z., Cingöz, O. & Goff, S. P. NP220 mediates silencing of unintegrated retroviral DNA. *Nature* **564**, 278–282 (2018).

10. 10.

Yurkovetskiy, L. et al. Primate immunodeficiency virus proteins Vpx and Vpr counteract transcriptional repression of proviruses by the HUSH complex. *Nat. Microbiol.* **3**, 1354–1361 (2018).

11. 11.

Greenwood, E. J. D. et al. Promiscuous targeting of cellular proteins by Vpr drives systems-level proteomic remodeling in HIV-1 infection. *Cell Rep.* **27**, 1579–1596.e7 (2019).

12. 12.

Chougui, G. et al. HIV-2/SIV viral protein X counteracts HUSH repressor complex. *Nat. Microbiol.* **3**, 891–897 (2018).

13. 13.

Tchasovnikarova, I. A. et al. Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. *Nat. Genet.* **49**, 1035–1044 (2017).

14. 14.

Douse, C. H. et al. Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. *Nat. Commun.* **9**, 651 (2018).

15. 15.

Douse, C. H. et al. TASOR is a pseudo-PARP that directs HUSH complex assembly and epigenetic transposon control. *Nat. Commun.* **11**, (2020).

16. 16.

Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).

17. 17.

Ciuffi, A. et al. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* **11**, 1287–1289 (2005).

18. 18.

Ding, S. et al. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473–483 (2005).

19. 19.

Scott, A. F. et al. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics* **1**, 113–125 (1987).

20. 20.

Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268–274 (2004).

21. 21.

Wang, G. Z., Wang, Y. & Goff, S. P. Histones are rapidly loaded onto unintegrated retroviral DNAs soon after nuclear entry. *Cell Host Microbe* **20**, 798–809 (2016).

22. 22.

Lacadie, S. A. & Rosbash, M. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. *Mol. Cell* **19**, 65–75 (2005).

23. 23.

Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24**, 363–367 (2000).

24. 24.

Keating, C. P. et al. The A-rich RNA sequences of HIV-1 pol are important for the synthesis of viral cDNA. *Nucleic Acids Res.* **37**, 945–956 (2009).

25. 25.

Bühler, M. RNA turnover and chromatin-dependent gene silencing. *Chromosoma* **118**, 141–151 (2009).

26. 26.

Kloc, A., Zaratiegui, M., Nora, E. & Martienssen, R. RNA interference guides histone modification during the S phase of chromosomal replication. *Curr. Biol.* **18**, 490–495 (2008).

27. 27.

Moazed, D. Mechanisms for the inheritance of chromatin states. *Cell* **146**, 510–518 (2011).

28. 28.

Busch, A. & Hertel, K. J. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA* **3**, 1–12 (2012).

29. 29.

Tilgner, H. et al. Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* **16**, 996–1001 (2009).

30. 30.

Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon–intron structure. *Nat. Struct. Mol. Biol.* **16**, 990–995 (2009).

31. 31.

Veloso, A. et al. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* **24**, 896–905 (2014).

32. 32.

Shaul, O. How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* **91**, 145–155 (2017).

33. 33.

Jacobs, F. M. J. et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242–245 (2014).

34. 34.

Schmidt, D. & Durrett, R. Adaptive evolution drives the diversification of zinc-finger binding domains. *Mol. Biol. Evol.* **21**, 2326–2339 (2004).

35. 35.

Piovesan, A., Caracausi, M., Antonaros, F., Pelleri, M. C. & Vitale, L. GeneBase 1.1: A tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. *Database* **2016**, baw153 (2016).

36. 36.

Gegonne, A. et al. The general transcription factor TAF7 is essential for embryonic development but not essential for the survival or differentiation of mature T cells. *Mol. Cell. Biol.* **32**, 1984–1997 (2012).

37. 37.

Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

38. 38.

Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

39. 39.

Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

40. 40.

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

41. 41.

Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

42. 42.

Smith, T., Heger, A. & Sudbery, I. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).

43. 43.

Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

44. 44.

Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

45. 45.

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).

46. 46.

McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).

47. 47.

Heger, A., Webber, C., Goodson, M., Ponting, C. P. & Lunter, G. GAT: A simulation framework for testing the association of genomic intervals. *Bioinformatics* **29**, 2046–2048 (2013).

48. 48.

Penzkofer, T., Dandekar, T. & Zemojtel, T. L1Base: From functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Res.* **33**, D498–D500 (2005).

49. 49.

Penzkofer, T. et al. L1Base 2: More retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Res.* **45**, D68–D73 (2017).

50. 50.

Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2006).

Acknowledgements

We thank the Genomics Core at Cancer Research UK Cambridge Institute for next-generation sequencing; JCBC and CIMR Core Facilities, particularly R. Schulte, A. P. Harrison and their teams for assistance with FACS; J. Cohen-Gold and E. Greenwood and other members of the Lehner lab for helpful discussions; J. Wysocka for L1-GFP and (opt)-L1-GFP plasmids; and S. Menzies for generation of TASOR and PPHLN HA KI cell lines. This work was supported by the Wellcome Trust Principal Research Fellowship to P.J.L. (210688/Z/18/Z). M.S. was supported by a Boehringer Ingelheim Fonds PhD fellowship (2017-2020). The work was further supported by the NIHR Cambridge Biomedical Research Centre.

Author information

Affiliations

1. Cambridge Institute for Therapeutic Immunology and Infectious Disease, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, UK

Marta Seczynska, Stuart Bloor & Paul J. Lehner

2. Data Sciences and Quantitative Biology, Discovery Sciences, AstraZeneca, Cambridge Biomedical Campus, Cambridge, UK

Sergio Martinez Cuesta

Contributions

M.S. designed, performed and analysed all experiments except from Northern blots and RT-PCR for GFP reporters which were conducted by S.B. S.M.C. performed all bioinformatics analyses with input from M.S. M.S. and P.J.L. conceived the study and wrote the manuscript.

Corresponding author

Correspondence to [Paul J. Lehner](#).

Ethics declarations

Competing interests

S.M.C. is an employee of AstraZeneca. All other authors declare no competing interests.

Peer review information

Nature thanks Todd Macfarlan and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

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

Extended data figures and tables

[Extended Data Fig. 1 HUSH repression of the L1 transgene is independent of the integration mode and site and is due to the L1 ORF2 sequence.](#)

a, Schematic of L1-iRFP reporter. The single mRNA transcript generates two proteins due to peptide bond skipping at the P2A sequence: the ORF1 (ORF1p) and iRFP (iRFPp). Changes in reporter transcription (e.g. due to H3K9me3-mediated silencing) affect iRFP expression. **b**, Western blot validating TASOR KO in HeLa cells with decreased levels of HUSH subunits periphilin and MPP8 due to TASOR depletion. β -actin is the loading control. **c,d**, Effect of reverse transcriptase inhibitor 3TC on the expression from L1 lentivirus (L1lenti) and L1 reporter integrated by transposase (L1pb). To validate that our reporters monitor expression only from the initial L1 integration, and no subsequent retrotransposition activity, we compared reporter expression in the presence and absence of reverse transcriptase inhibitor 3TC, which prevents retrotransposition and new L1 insertions. If expression from new insertions contributes to the iRFP signal, 3TC should decrease iRFP, in particular at time points ≥ 3 days. No such decrease in iRFP signal is observed with L1lenti and L1pb upon 3TC treatment, demonstrating that both reporters monitor expression from the initial L1 integration. As lentiviral integration requires reverse transcription, the 3TC was added 12h post transduction when reverse transcription will be complete. **e**, Flow cytometry histograms of expression of L1 lentivirus upon 50 μ M 3TC treatment in WT

or TASOR KO cells (left). 3TC was added 12 h post transduction and expression measured at day 5 post transduction. As a positive control, 3TC was used at the time of transduction (c – right hand panel). The absence of iRFP signal confirms inhibition of RT activity. Quantification of expression using geometric mean fluorescence intensity (gMFI) (right). **d**, Flow cytometry histograms of expression of L1pb reporter in the absence or presence of 50 μ M 3TC after 5 days of dox induction (left). Quantification of expression using gMFI (right). **e**, Establishment of repression of L1 reporter lentivirus in WT and HUSH KOs: TASOR, MPP8 or periphilin KO HeLa cells, and KOs of HUSH-effectors MORC2 and SETDB1. HUSH/HUSH effector KO (all GFP+) were mixed with WT cells and transduced with the L1-iRFP reporter at high multiplicity of infection (MOI). Expression of the reporter was measured by flow cytometry 72h post transduction. Western blot validating KOs with p97 as a loading control (bottom panel). * marks non-specific band. **f**, Schematic of lentiviral vector driven by dox-responsive promoter for expression of Human Immunodeficiency virus 2 (HIV-2) viral protein X (Vpx) to induce TASOR depletion (left panel). Western blot validating TASOR depletion 6 days after Vpx induction by dox with tubulin as a loading control (right panel). Flow cytometry histogram showing expression of integrated L1 lentivirus reporter before and after TASOR depletion by Vpx (bottom panel). **g**, Expression of L1 reporter lentivirus in: WT HeLa (WT+reporter) (grey histogram), HeLa cells in which TASOR was depleted after the integration of L1 lentivirus (+TASOR KD) (purple histogram) and re-expression of mCherry-TASOR (+TASOR KD +mCherry-TASOR) (grey, dotted histogram). This experiment was repeated in Jurkat cells with similar results. **h**, Northern blot showing increased mRNA from L1pb reporter in TASOR KO cells 24 h post dox induction using iRFP probe. Full-length L1pb mRNA is the predominant RNA produced from the L1pb reporter. **i**, The expression of reporters with ORF2 or ORF1 sequences placed downstream of GFP in WT or TASOR-depleted (TASOR KD; see Extended Data Fig. [2a](#)) HeLa cells measured by flow cytometry (upper panel). HUSH-mediated repression of reporters with ORF2 or ORF1 sequences placed upstream of GFP. CRISPR–Cas9 of TASOR (TASOR KO) after reporter integration (bottom panel). **j**, Northern blot analysis of mRNAs produced from L1pb reporter in which ORF2 (4kb) sequence was replaced by 4 tandem repeats of ORF1 (1kb). RNA was isolated from the mix of WT and TASOR KO cells and iRFP probe was used to detect reporter mRNA (Fig. [1d](#)).

Extended Data Fig. 2 HUSH-mediated repression of cDNAs and ORF2 fragments correlates with length of the transgene.

a, Western blot showing TASOR depletion in TASOR KD HeLa cells. **b**, Schematic of assay for the establishment of silencing of lentiviral transgenes. **c**, Schematic of the gating strategy in ‘one pot’ assay for establishment of silencing. mCherry+ WT and mCherry- TASOR KO HeLa cells were defined based on the mCherry signal and the GFP signal for each of these subpopulations is subsequently plotted on the histogram.

d, HUSH-mediated repression of the lentivirus encoding fusion of endonuclease dead Cas9 and KRAB domain (dCas9-KRAB) in HeLa (left) or Jurkat cells (right) measured by flow cytometry. mCherry fluorescence reports mRNA levels from the reporter. For Jurkat cells, a sgRNA targeting the TSS of *TASOR* was used to deplete TASOR. **e**, HUSH-mediated repression monitored at different time points post infection and after selection with the antibiotic for the transgene-delivered antibiotic resistance gene. **f** HUSH-mediated repression monitored 48h after transduction of HeLa WT and TASOR KD with lentiviral reporter at different range of MOI. **e** and **f** were repeated with different reporters with similar results. **g**, Scatter plot illustrating a significant correlation between HUSH-mediated repression and length of the insert sequence in the GFP reporters. Each point represents a reporter with different cDNA sequence. Pearson correlation $r = 0.7115$, two-sided $p = 0.0003$; 95% CI [0.40 to 0.87] **h**, Expression of GFP non-coding lentiviral reporters bearing different short cDNA sequences in WT and TASOR KD HeLa cells measured by flow cytometry 72h post transduction. **i**, HUSH-mediated repression of GFP bearing the indicated untranslated ORF2 fragments measured by flow cytometry. **j**, Quantification of the HUSH-mediated repression of GFP untranslated reporters bearing full length ORF2 or ORF2 fragments, $n = 3$ biological replicates \pm SD (left). **k**, RT-PCR analysis of transcripts from GFP reporters bearing ORF2 fragments with primers flanking ORF2 fragments (right). Product sizes corresponding to full length transcripts are 1.7 kb and 3.8 kb for reporters with 1-4 fragments and Δ 1-4 fragments respectively.

[Source data](#)

[Extended Data Fig. 3 Susceptibility to HUSH-repression is governed by high adenine content in the sense strand and transgene length.](#)

a, Scatter plot illustrating the relationship between HUSH-mediated repression and AT content of the insert sequence in the GFP reporter. Each point represents a reporter with different cDNA sequence. Reporters were assigned into three groups according to the length of the insert cDNA sequence (orange, green and grey) and Pearson r correlation was quantified for each group. **b**, Expression of GFP reporter bearing untranslated sequence of native or codon-optimized ORF2 (with increased GC content) in WT and TASOR KD HeLa cells measured by flow cytometry. Quantification of $n = 3$ independent experiments in ‘e’. **c**, Scatter plots illustrating the relationship between HUSH-mediated repression and nucleotide content of the insert sequence in the GFP reporter (left) with the significance of the two-sided Pearson correlation between HUSH-mediated repression and nucleotide content (right). Dotted line on the graph corresponds to p -value = 0.05; for exact p -value see source data. **d**, HUSH-mediated repression of GFP lentiviral reporters bearing native ORF2 (A-rich) or reverse-complement ORF2 sequence (T-rich) measured 4 days post transduction (right). To prevent premature transcription termination, two putative polyadenylation

sites were deleted from ORF2 sequence (AATAAA at position 228-233 of reverse complement ORF2 and ATTAAA at 123-129). Relative contribution of A and T nucleotides to the nucleotide content of the insert (left). **e**, Quantification of the HUSH-mediated repression of GFP untranslated reporters bearing native, codon-optimized (codon opt.) or reverse complement (reverse compl.) ORF2; n = 4 (native and codon opt.) and n = 2 (for reverse compl.) biological replicates \pm SD, normalized to gMFI of native ORF2 reporter in WT cells. **f**, The expression of transposase-integrated reporters bearing ORF2 or reverse complement ORF2 sequence in WT or TASOR KO HeLa cells measured by flow cytometry. **g**, Analysis of mRNA produced from reporters in **f**, by Northern blot. RNA was isolated from the mix of WT and TASOR KO cells. **h**, Quantification of HUSH-mediated repression of L1pb reporter in which ORF2 sequence was replaced by 1 ORF1, 3 ORF1 or 4 ORF1 tandem repeats. mean of n = 3 biological replicates \pm SD ***p < 0.001 (for exact p value see source data); one-way ANOVA post-hoc pairwise comparisons with Bonferroni correction.

[Source data](#)

Extended Data Fig. 4 HUSH represses non-integrated DNAs and requires transcription to maintain repression.

a, Transduction of cells with lentiviral reporter in the presence or absence of the integrase inhibitor raltegravir to test the establishment of reporter silencing in the presence or absence of reporter integration. **b**, Representative flow cytometry histograms of expression from integrated or unintegrated lentiviral reporter in WT or TASOR KD HeLa cells. As unintegrated lentivirus is poorly expressed, the reporter with the synthetic ORF2 sequence was used since it provides higher expression than the native ORF2 sequence. **c**, Western blot showing CRISPR/Cas9 mediated depletion of TASOR in the population of 293T cells (TASOR KD). β -actin is a loading control. **d**, Flow cytometry histograms of expression from pcDNA3.1 plasmid transfected into WT or TASOR KD 293T cells. In contrast to lentivirus or plasmids for piggyBac-mediated integration, pcDNA3.1 lacks terminal repeats (ITRs). **e**, RT-qPCR quantifying transcript levels from SFFV-driven or promoter-less L1 lentiviral reporter integrated into WT and TASOR KO cells. Normalized to WT with SFFV-driven L1. n = 3 biological replicates (independent polyclonal integrations of the reporters) \pm SD **f**, Northern blot analysis of mRNA produced from SFFV-driven or promoter-less L1 lentiviral reporter in WT and TASOR KO cells. **g**, ChIP-qPCR quantifying H3K9me3 at promoter-less L1 lentiviral reporter in clonal WT HeLa populations normalized to polyclonal WT population with SFFV-driven L1. n = 6 biological replicates \pm SD, *p = 0.03 one-sample Wilcoxon test **h**, Genome browser track depicting H3K9me3 ChIPseq signal over control, HUSH-repressed locus in WT and TASOR KO HeLa cells harbouring SFFV-driven or promoter-less L1 reporter - related to Fig. 2b. **i**, CRISPR/Cas9-mediated deletion of the *TAF7* promoter region (schematic, upper left)

reduces *TAF7* transcription measured by RT-qPCR and normalized to WT (bottom right). ChIP-qPCR quantifying H3K9me3 and total H3 at the locus (bottom left). n = 2 biological replicates x 3 independent experiments \pm SD; **p = 0.0023, ##p = 0.009 one-way ANOVA post-hoc pairwise comparisons vs WT with Bonferroni correction. Cas9-cleavage sites indicated by scissors, green arrows indicate primers used to validate the deletion by genomic PCR and red arrows indicate position of the primers used in ChIP-qPCR. Gel image (upper right) confirms promoter deletion. **j**, Genome browser track depicting H3K9me3 ChIPseq signal over *TAF7* locus or control HUSH-repressed locus in WT, TASOR KO HeLa and HeLa with deletion of *TAF7* promoter. **k**, Western blot of HeLa cells with HA knocked into endogenous locus of *TASOR* or *PPHLN1*. **i**, Schematic of HUSH-sensitive and HUSH-resistant reporter constructs (upper schematic). Expression from the reporters is driven by dox-responsive promoter. Human beta globin coding sequence (*HBBcds*), instead of ORF1 as in the standard L1 reporter, is followed by P2A-iRFP and, for the HUSH sensitive reporter, by ORF2 sequence. HUSH-sensitive and HUSH-resistant reporters were integrated into control, HA-TASOR, PPHLN1-HA cell lines - resulting in six independent cell lines in total. In cell lines with the HUSH-sensitive reporter, SETDB1 function was then disrupted by CRISPR/Cas9-mediated knockout and mixed, polyclonal KO populations were used for RIP-qPCR. Flow cytometry histograms of expression from HUSH-sensitive or HUSH-resistant reporter in HA KI and control cell lines 48h after induction with dox (bottom). For the HUSH-sensitive reporter the expression is shown in WT and SETDB1 KO cells. Right panel: Relative levels of transcripts from reporters for RIP-qPCR (in SETDB1 KO) in nuclear fraction normalized to *ACTB*; n = 2 technical replicates **m**, Validation of SETDB1 depletion by CRISPR/Cas9 in TASOR or PPHLN1 HA-KI cells by western blot. β -actin as loading control. * marks non-specific band.

Extended Data Fig. 5 Periphilin specifically binds transcripts from evolutionary young, full-length L1 elements in WT and SETDB1-depleted cells.

a, Sequencing tracks showing insertion of sequence of HA-tag (marked as dashed box) into *PPHLN1* locus. Underlined is the stop codon. Nucleotide substitutions to make modified locus sgRNA-resistant are marked as small letters. **b**, Western blot showing SETDB1 depletion in SETDB1 KO (mix) Periphilin-HA HEK293Ts and control HEK293Ts. p97/VCP as a loading. **c,d**, Enrichment of periphilin RIP-seq peaks at different repetitive elements in **c**, WT cells and **d**, WT (left) and SETDB1 KO (mix) (right). Significant enrichment is defined as a fold change score greater than one with Benjamini–Hochberg empirical adjusted one-sided p-value calculated using simulations and genomic association testing⁴⁷, ***q < 0.001 (for exact p-values see source data). **e**, Fraction of full length, non-full length L1s and L1s from different families overlapping with periphilin RIPseq peaks. Full length L1s definitions are

based on L1Base^{48,49}. Blue heatmap indicates age of L1 families predicted from the phylogenetic analysis⁵⁰. Periphilin-bound L1Hs may be underestimated in comparison to L1PA2-L1PA3 due to lower mappability of L1Hs as this is the least sequence-divergent L1 family. **f**, Genome browser tracks showing periphilin RIP signal over intronic L1s, Tigger DNA transposon and 3'UTR of *ZNF37A*

[Source data.](#)

Extended Data Fig. 6 Introns protect different reporters from HUSH and are continuously required to prevent repression.

a, Quantification of H3K9me3 and RNAseq signal over endogenous genes in WT and TASOR KO K562 cells from a publicly available dataset². None of these endogenous genes are HUSH-repressed, unlike lentiviral reporters containing cDNA sequences of these genes. **b**, Northern blot analysis of mRNAs produced from intronless reporter or reporter with *HBB* IVS2 cloned within the iRFP gene. ACTB is a loading control. **c**, Flow cytometry histograms showing expression from GFP and GFP-ORF2 intronless or intron-containing lentiviral reporters in WT and TASOR KD HeLa cells 72h post transduction (bottom). Schematic of the construct (top). To prevent intron splicing during transcription in the virus-producing cells, the reporter cassette driven by the SFFV promoter was cloned in reverse orientation with respect to lentiviral transcription. The polyadenylation signal (pA) in reverse orientation provides a signal for termination of transcription from the reporter cassette in transduced cells. ORF2 is untranslated and intron (*HBB* IVS2) is cloned 5' of ORF2. SA-splice acceptor, SD-splice donor. **d**, HUSH-mediated repression of integrated intronless or intron-containing ORF2 piggyBac reporters measured by flow cytometry (histograms in centre panel) and calculated as the ratio of reporter expression in TASOR KO and WT HeLa (right). Expression from the reporter is driven by a dox-responsive CMV promoter. Reporters contain either human beta globin coding sequence or genomic sequence (containing 2 introns) followed by P2A-iRFP and ORF2 sequences (schematics on the left). A HUSH-resistant reporter without ORF2 is the negative control. n biological replicates (independent polyclonal integrations of the reporters) ± SD; ***p ≤ 0.0001, one-way ANOVA post-hoc pairwise comparisons vs –introns with Bonferroni correction. **e**, HUSH-mediated repression of integrated intronless or intron-containing Cas9 piggyBac reporters measured by flow cytometry (histograms in centre panel) and calculated as the ratio of reporter expression in TASOR KO and WT HeLa (right). Expression from the reporter is driven by a dox-responsive CMV promoter. Reporters contain either *HBB* coding sequence or genomic sequence (containing 2 introns) followed by P2A-iRFP and Cas9 sequences (schematics on left panel). A HUSH-resistant reporter without Cas9 is the negative control. n biological replicates (independent polyclonal integrations of the reporters) ± SD; ***p ≤ 0.0001, one-way ANOVA post-hoc pairwise comparisons vs –introns with Bonferroni correction. **f**,

HUSH-mediated repression of reporter with intron removed by Cre-loxP recombination following the reporter integration (upper schematic). Flow cytometry histograms of expression from iRFP-ORF2 reporters driven by EF1a promoter: (i) intronless or (ii) reporter-bearing intron (*HBB* IVS2) flanked by loxP sites in the absence or presence of Cre expression (left). Gel image (right) confirms intron deletion.

Extended Data Fig. 7 Intron insertion reduces HUSH-mediated repression and Periphilin binding to reporter transcripts.

a, Representative flow cytometry histograms of expression from reporters in Fig. 3c in WT and TASOR KO HeLa cells. The 5' and 3' control (asGFP) is the antisense GFP 'stuffer' sequence **b**, ChIP-qPCR quantifying H3K9me3 and total H3 levels at intronless or reporter with introns (*HBB* IVS2) inserted at 5' and 3' of ORF2 (from Fig. 3c and Extended Data Fig. 7a). n = 4 independent experiments ± SD; ***p = 0.0003 and ###p = 0.0006 vs -intron WT, ratio paired two-tailed t-test. **c**, Gel images confirming splicing of introns at 5' and 3' of ORF2 from iRFP-ORF2 reporter transcripts (from Fig. 3c and Extended Data Fig. 7a) by PCR. **d**, Northern blot analysis of mRNAs produced from reporters with intron or control sequence inserted 5' and 3' of ORF2 (~5kb), or HUSH-resistant reporter without ORF2 (~1.5kb). RNA was isolated from the mix of WT and TASOR KO cells. **e**, PCR analysis of splicing of different introns 5' of ORF2 from iRFP-ORF2 reporter transcripts (from Fig. 3d and Extended Data Fig. 7f). **f**, Representative flow cytometry histograms of expression from iRFP-ORF2 reporter with introns from *ACTB* (0.4kb) or a short, chimeric intron (0.13kb) cloned 5' of ORF2 in WT and TASOR KO HeLa cells. Experiment repeated independently with similar results; quantification of n = 3-4 biological replicates in Fig. 3f. **g**, Schematic of intronless and intron-containing reporter constructs for periphilin RIP-qPCR (upper schematic). Reporters were integrated into WT 293T or periphilin-HA 293Ts - resulting in four independent cell lines. SETDB1 function was disrupted by CRISPR/Cas9-mediated knockout and mixed, polyclonal KO populations were used for RIP-qPCR. Flow cytometry histograms of expression from reporters in PPHLN-HA and control cell lines 48h after induction with dox (bottom). **h**, Validation of SETDB1 depletion by CRISPR/Cas9 in PPHLN1 HA-KI cells by western blot. β-actin as loading control. * marks non-specific band. **i**, Relative levels of transcripts from reporters for RIP-qPCR (in SETDB1 KO) in nuclear fraction normalized to *ACTB*; n = 2 technical replicates. **j**, RIP-qPCR showing decreased association of periphilin with RNA from intron-containing reporter. L1Hs and *ACTB* RNA are a positive and negative control, respectively. Data are mean ± SD; n = 3 independent experiments; and normalized to input. ***p = 0.0009 vs -intron, one-way ANOVA post-hoc pairwise comparison with Bonferroni correction.

[Extended Data Fig. 8 Sequences engineered for efficient splicing do not protect against HUSH repression.](#)

a, Schematic of intron mutations in reporters from Fig. 3e. **b**, Analysis of splicing of mutant introns inserted 5' of ORF2 from iRFP-ORF2 reporter transcripts by PCR (from Fig. 3e). **c**, Representative flow cytometry histograms of expression from reporters containing spliced stuffer sequences in WT and TASOR KO HeLa cells.

[Extended Data Fig. 9 Transcribed processed pseudogenes and protein coding retrogenes, but not their parent genes, are bound and silenced by HUSH.](#)

Genome browser tracks showing HUSH-dependent H3K9me3, HUSH/MORC2-occupancy and RNA-seq in WT and HUSH KO K562 cells at: **a**, additional, representative loci of retrogenes; **b**, *FNBPIP1* pseudogene (left) and its parent gene *FNBPI* (right); **c**, *UTP14C* retrogene (left) and its parent gene *UTP14A* (right) **d**, at the locus of *MAB21L2*, a non-transcribed paralog of HUSH-repressed *MAB21L1* retrogene, Data from 2. **e**, HUSH-repressed genes obtained from the dataset in ref. 2: 378 genes were obtained when there was at least 30% reduction of H3K9me3 signal in all 3 knockout cell lines: TASOR KO, MPP8 KO and MORC2 KO (\log_2 FC H3K9me3 TASOR KO/WT ≤ -0.5 ; FDR significance ≤ 0.05 ; determined after a comparative assessment of counts between conditions ($n = 2$) using negative binomial generalized linear models as implemented in edgeR and corrected for multiple comparisons using FDR method). 104 of them were ZNF genes (including 8 ZNF pseudogenes) and the rest were inspected in IGV to determine the most probable reason for HUSH-repression e.g. overlap with L1 elements or other HUSH targets, pseudogene, retrogene, genes with signal over long exons, novel transcripts or antisense lncRNAs or loci with bidirectional promoter. The 11 resulting genes remained unannotated, either excluded because of low, background H3K9me3 over region or mapping artifacts or the reason for repression was unclear. Fraction of pseudogenes and retrogenes transcribed (average RNA-seq signal of all samples above > 0.1 RPKM or within transcriptionally active gene) (right) **f**, Genome browser tracks showing HUSH-dependent H3K9me3, HUSH/MORC2-occupancy and RNA-seq in WT and HUSH KO K562 cells at representative long exons. Data from ref. 2.

[Source data](#)

[Extended Data Fig. 10 Periphilin specifically binds to transcripts from intronless genomic loci.](#)

Genome browser tracks showing periphilin RIP signal over representative loci of processed pseudogenes **a**, and intronless genes **b**, in WT and SETDB1 KO (mix) HEK293T cells. **c**, Heatmap showing periphilin and control RIP signal (RPKM) over selected pseudogenes and their corresponding parent intron-containing genes. For *DUXA* and *AGGF1* parent genes, two (*DUXAP9*, *DUXAP10*) and four (*AGGF1P1*, *AGGF1P2*, *AGGF1P3*, *AGGF1P10*) pseudogenes are depicted. Data from periphilin RIPseq in SETDB1 KO (mix) cells (median of n = 4 independent experiments). **d**, Metagene profile of fold change of periphilin and control mean RIP-seq signal over three categories of genes: processed pseudogenes, intronless genes and intron-containing protein-coding genes. Only genes with periphilin RIPseq signal greater than 0.3 RPKM are considered (in each four RIP replicates in SETDB1 KO (mix) and two replicates in WT 293Ts). Genes where the periphilin signal enrichment peaks overlap with L1 elements are excluded. TSS-transcription start site, TTS-transcription termination site. Intronless protein-coding genes produce only intronless isoforms. **e**, Genome browser track showing periphilin RIPseq signal over representative locus of intron-containing gene (*BOD1L1*) with a long exon in WT and SETDB1 KO (mix) HEK293Ts.

[Source data](#)

Supplementary information

[Supplementary Information](#)

This file contains Supplementary Figs. 1, 2.

[Reporting Summary](#)

[Peer Review File](#)

[Supplementary Table 1](#)

Resource table: Plasmids, oligonucleotides, sgRNAs and sequences used in the study.

Source data

[Source Data Fig. 2](#)

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

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

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

[**Source Data Extended Data Fig. 9**](#)

[**Source Data Extended Data Fig. 10**](#)

Rights and permissions

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

[**Reprints and Permissions**](#)

About this article

Cite this article

Seczynska, M., Bloor, S., Cuesta, S.M. *et al.* Genome surveillance by HUSH-mediated silencing of intronless mobile elements. *Nature* **601**, 440–445 (2022).
<https://doi.org/10.1038/s41586-021-04228-1>

- Received: 03 February 2021
- Accepted: 10 November 2021
- Published: 18 November 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04228-1>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04228-1>

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

- Article
- [Published: 22 December 2021](#)

MicroRNA sequence codes for small extracellular vesicle release and cellular retention

- [Ruben Garcia-Martin](#) [ORCID: orcid.org/0000-0003-2049-4960¹](#),
- [Guoxiao Wang¹](#),
- [Bruna B. Brandão](#) [ORCID: orcid.org/0000-0001-8762-6310¹](#),
- [Tamires M. Zanotto¹](#),
- [Samah Shah²](#),
- [Sandip Kumar Patel](#) [ORCID: orcid.org/0000-0002-0651-4438²](#),
- [Birgit Schilling](#) [ORCID: orcid.org/0000-0001-9907-2749²](#) &
- [C. Ronald Kahn](#) [ORCID: orcid.org/0000-0002-7583-9228¹](#)

[Nature](#) volume 601, pages 446–451 (2022)

- 18k Accesses
- 2 Citations
- 222 Altmetric
- [Metrics details](#)

Subjects

- [Metabolism](#)
- [miRNAs](#)

Abstract

Exosomes and other small extracellular vesicles (sEVs) provide a unique mode of cell-to-cell communication in which microRNAs (miRNAs) produced and released from one cell are taken up by cells at a distance where they can enact changes in gene expression^{1,2,3}. However, the mechanism by which miRNAs are sorted into exosomes/sEVs or retained in cells remains largely unknown. Here we demonstrate that miRNAs possess sorting sequences that determine their secretion in sEVs (EXOmotifs) or cellular retention (CELLmotifs) and that different cell types, including white and brown adipocytes, endothelium, liver and muscle, make preferential use of specific sorting sequences, thus defining the sEV miRNA profile of that cell type. Insertion or deletion of these CELLmotifs or EXOmotifs in a miRNA increases or decreases retention in the cell of production or secretion into exosomes/sEVs. Two RNA-binding proteins, Alyref and Fus, are involved in the export of miRNAs carrying one of the strongest EXOmotifs, CGGGAG. Increased miRNA delivery mediated by EXOmotifs leads to enhanced inhibition of target genes in distant cells. Thus, this miRNA code not only provides important insights that link circulating exosomal miRNAs to tissues of origin, but also provides an approach for improved targeting in RNA-mediated therapies.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Cell-type-specific miRNAs in sEVs and cells and selectivity of sEV versus cellular distribution for certain miRNAs.

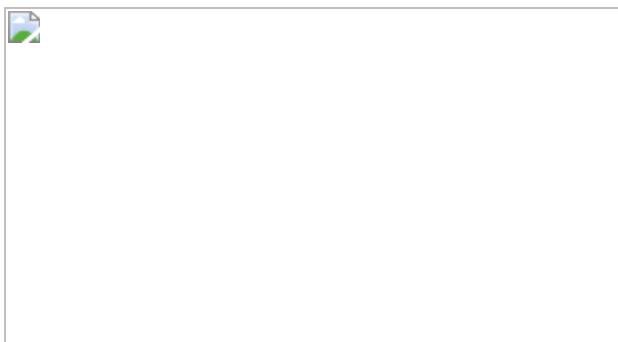


Fig. 2: Motifs over-represented in miRNAs preferentially sorted into exosomes/sEVs (EXOmotifs) or retained in cells (CELLmotifs) for each

cell type.



Fig. 3: CELLmotifs and EXOmotifs regulate miRNA distribution, and Alyref and Fus participate in sorting of miRNAs containing CNGGNG-type EXOmotifs.

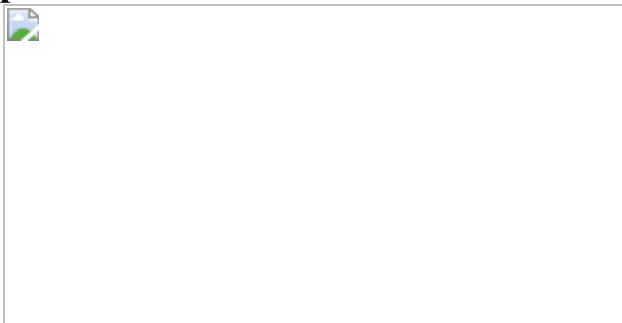
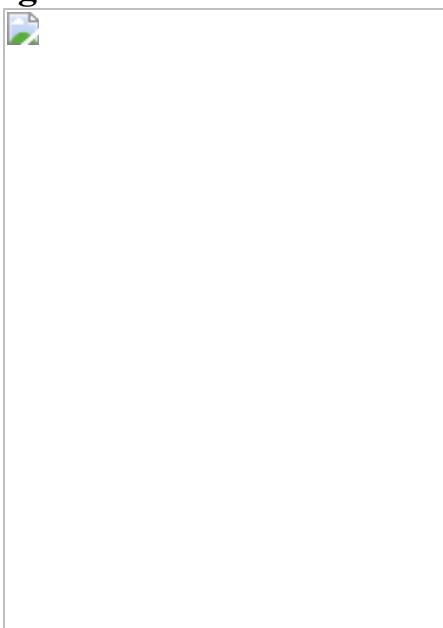


Fig. 4: Incorporation of EXOmotifs enhances miRNA delivery and target inhibition.



Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its [Supplementary Information](#). The miRNA profiling raw dataset can be found in Supplementary Table [1](#) as negative ΔC_t values normalized by the average C_t of each corresponding sample. Average normalized raw values for each cell type and the statistics for each miRNA in cell bodies and exosomes/sEVs are shown in Supplementary Tables [2](#) and [4](#), respectively. Significance tables for comparisons between each cell type and the other four cell types in cell bodies and sEVs can be found in Supplementary Tables [3](#) and [5](#), respectively. An sEV and cellular enrichment table for each miRNA can be found in Supplementary Table [6](#). The primary hepatocyte miRNA profile is shown in Supplementary Table [7](#). The miRNA profile from the ultracentrifugation + size-exclusion chromatography experiment is shown in Supplementary Table [8](#). smRNA-seq data can be found in Supplementary Table [10](#). Pre-miRNA sequences of the expressed miRNAs used in this study can be found in Supplementary Table [11](#), and their mature forms are displayed in Extended Data Figs. [7–9](#). The Ensembl and miRBase databases are publicly available at <http://www.ensembl.org> and <http://www.mirbase.org>, respectively.

Experimental procedures have also been uploaded to EV-TRACK^{[42](#)} with reference nos EV200052 and EV210287. The mass spectrometry analysis can be found in Supplementary Table [9](#), and the raw data have been deposited with MassIVE ID MSV000086780 and are also available at ProteomeXchange with ID PXD023895. Additional mass spectrometry details from DIA and DDA acquisitions, such as protein identification and quantification details, are available at the repositories (including all generated Spectronaut and ProteinPilot search engine files).

References

1. 1.

Thomou, T. et al. Adipose-derived circulating miRNAs regulate gene expression in other tissues. *Nature* **542**, 450–455 (2017).

2. 2.

Valadi, H. et al. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* **9**, 654–659 (2007).

3. 3.

Skog, J. et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat. Cell Biol.* **10**, 1470–1476 (2008).

4. 4.

Mori, M. A., Ludwig, R. G., Garcia-Martin, R., Brandao, B. B. & Kahn, C. R. Extracellular miRNAs: from biomarkers to mediators of physiology and disease. *Cell Metab.* **30**, 656–673 (2019).

5. 5.

Tosar, J. P. et al. Assessment of small RNA sorting into different extracellular fractions revealed by high-throughput sequencing of breast cell lines. *Nucleic Acids Res.* **43**, 5601–5616 (2015).

6. 6.

Arroyo, J. D. et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl Acad. Sci. USA* **108**, 5003–5008 (2011).

7. 7.

Hunter, M. P. et al. Detection of microRNA expression in human peripheral blood microvesicles. *PLoS ONE* **3**, e3694 (2008).

8. 8.

Crewe, C. et al. An endothelial-to-adipocyte extracellular vesicle axis governed by metabolic state. *Cell* **175**, 695–708 (2018).

9. 9.

Zomer, A. et al. In vivo imaging reveals extracellular vesicle-mediated phenocopying of metastatic behavior. *Cell* **161**, 1046–1057 (2015).

10. 10.

Abels, E. R. et al. Glioblastoma-associated microglia reprogramming is mediated by functional transfer of extracellular miR-21. *Cell Rep.* **28**, 3105–3119 (2019).

11. 11.

Kosaka, N. et al. Secretory mechanisms and intercellular transfer of microRNAs in living cells. *J. Biol. Chem.* **285**, 17442–17452 (2010).

12. 12.

Shurtleff, M. J., Temoche-Diaz, M. M., Karfiliis, K. V., Ri, S. & Schekman, R. Y-box protein 1 is required to sort microRNAs into exosomes in cells and in a cell-free reaction. *eLife* **5**, e19276 (2016).

13. 13.

Wei, Z. et al. Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. *Nat. Commun.* **8**, 1145 (2017).

14. 14.

Cha, D. J. et al. KRAS-dependent sorting of miRNA to exosomes. *eLife* **4**, e07197 (2015).

15. 15.

Jeppesen, D. K. et al. Reassessment of exosome composition. *Cell* **177**, 428–445 (2019).

16. 16.

Santangelo, L. et al. The RNA-binding protein SYNCRIPI is a component of the hepatocyte exosomal machinery controlling

microRNA sorting. *Cell Rep.* **17**, 799–808 (2016).

17. 17.

Temoche-Diaz, M. M. et al. Distinct mechanisms of microRNA sorting into cancer cell-derived extracellular vesicle subtypes. *eLife* **8**, e47544 (2019).

18. 18.

Villarroya-Beltri, C. et al. Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. *Nat. Commun.* **4**, 2980 (2013).

19. 19.

Thery, C. et al. Minimal Information for Studies of Extracellular Vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J. Extracell. Vesicles* **7**, 1535750 (2018).

20. 20.

Lagos-Quintana, M. et al. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).

21. 21.

Jopling, C. Liver-specific microRNA-122: biogenesis and function. *RNA Biol.* **9**, 137–142 (2012).

22. 22.

Horaka, M. N. J. & Bienertova-Vaskuac, J. Muscle-specific microRNAs in skeletal muscle development. *Dev. Biol.* **410**, 1–13 (2016).

23. 23.

Gao, T., Shu, J. & Cui, J. A systematic approach to RNA-associated motif discovery. *BMC Genomics* **19**, 146 (2018).

24. 24.

Bartel, D. P. Metazoan microRNAs. *Cell* **173**, 20–51 (2018).

25. 25.

Statello, L. et al. Identification of RNA-binding proteins in exosomes capable of interacting with different types of RNA: RBP-facilitated transport of RNAs into exosomes. *PLoS ONE* **13**, e0195969 (2018).

26. 26.

Shi, M. et al. ALYREF mainly binds to the 5' and the 3' regions of the mRNA in vivo. *Nucleic Acids Res.* **45**, 9640–9653 (2017).

27. 27.

Zhang, T. et al. FUS regulates activity of microRNA-mediated gene silencing. *Mol. Cell* **69**, 787–801 (2018).

28. 28.

Bae, Y. et al. miRNA-34c regulates Notch signaling during bone development. *Hum. Mol. Genet.* **21**, 2991–3000 (2012).

29. 29.

Hu, S. et al. MicroRNA-34c downregulation ameliorates amyloid- β -induced synaptic failure and memory deficits by targeting VAMP2. *J. Alzheimers Dis.* **48**, 673–686 (2015).

30. 30.

Ngalame, N. N., Tokar, E. J., Person, R. J., Xu, Y. & Waalkes, M. P. Aberrant microRNA expression likely controls *RAS* oncogene

activation during malignant transformation of human prostate epithelial and stem cells by arsenic. *Toxicol. Sci.* **138**, 268–277 (2014).

31. 31.

Paraskevopoulou, M. D. et al. DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res.* **41**, D239–D245 (2013).

32. 32.

Chevillet, J. R. et al. Quantitative and stoichiometric analysis of the microRNA content of exosomes. *Proc. Natl Acad. Sci. USA* **111**, 14888–14893 (2014).

33. 33.

Fasshauer, M. et al. Essential role of insulin receptor substrate-2 in insulin stimulation of Glut4 translocation and glucose uptake in brown adipocytes. *J. Biol. Chem.* **275**, 25494–25501 (2000).

34. 34.

Thery, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr. Protoc. Cell Biol.* **Chapter 3**, Unit 3.22, (2006).

35. 35.

Peltier, H. J. & Latham, G. J. Normalization of microRNA expression levels in quantitative RT–PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA* **14**, 844–852 (2008).

36. 36.

Collins, B. C. et al. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass

spectrometry. *Nat. Commun.* **8**, 291 (2017).

37. 37.

Schilling, B., Gibson, B. W. & Hunter, C. L. Generation of high-quality SWATH(^R) acquisition data for label-free quantitative proteomics studies using tripleTOF(^R) mass spectrometers. *Methods Mol. Biol.* **1550**, 223–233 (2017).

38. 38.

Ritchie, M. E. et al. Empirical array quality weights in the analysis of microarray data. *BMC Bioinf.* **7**, 261 (2006).

39. 39.

Burger, T. Gentle introduction to the statistical foundations of false discovery rate in quantitative proteomics. *J. Proteome Res.* **17**, 12–22 (2018).

40. 40.

Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

41. 41.

Kullback, S. & Leibler, R. A., On information and sufficiency, *Ann. Math. Stat.* **22**, 79–86 (1951).

42. 42.

Consortium, E.-T. et al. EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research. *Nat. Methods* **14**, 228–232 (2017).

Acknowledgements

This work was supported by an NIH grant (R01DK082659 to C.R.K.). R.G.-M. was supported by a Deutsche Forschungsgemeinschaft fellowship (GA 2426/1-1, Germany). G.W. was supported by a T32 grant (DK007260). T.M.Z. was supported by a FAPESP fellowship (2017/10179-9, Brazil). We also received support from the Joslin DRC (P30DK036836). S.S., S.K.P. and B.S. were supported by Shared Instrumentation Grant S10 OD016281 to the Buck Institute. S.K.P. was supported by a fellowship from the Glenn Foundation. We thank A.M. Ciulla and the BPF Genomics Core Facility at Harvard Medical School for their expertise and instrument availability that supported this work. We thank M. Ericsson from the Electron Microscopy Facility at the Harvard Medical School for assistance with electron microscopy experiments. We thank J.M. Dreyfuss and H. Pan from the Joslin Bioinformatics and Biostatistics Core for assistance with data analysis and A. Wood from the Joslin Flow Cytometry Core for assistance with experiments.

Author information

Affiliations

1. Section of Integrative Physiology and Metabolism, Joslin Diabetes Center, Harvard Medical School, Boston, MA, USA

Ruben Garcia-Martin, Guoxiao Wang, Bruna B. Brandão, Tamires M. Zanotto & C. Ronald Kahn

2. The Buck Institute for Research on Aging, Novato, CA, USA

Samah Shah, Sandip Kumar Patel & Birgit Schilling

Contributions

R.G.-M. designed research, performed experiments, analysed the data shown in Figs. 1–4 and Extended Data Figs. 1–10 and wrote the manuscript. G.W. helped with vector generation for the miRNA overexpression shown in Fig. 3a–f and Extended Data Figs. 7–9. B.B.B. helped with the size-exclusion chromatography experiments shown in

Extended Data Fig. 6. T.M.Z. helped with analysis of the miRNA profiling and motifs shown in Figs. 1 and 2. S.S., S.K.P. and B.S. performed the proteomic study and its analysis shown in Fig. 3g–j and Extended Data Fig. 10a–c. C.R.K. designed the research, wrote the manuscript and supervised the project.

Corresponding author

Correspondence to [C. Ronald Kahn](#).

Ethics declarations

Competing interests

A patent submitted by R.G.-M. and C.R.K. (US patent no. US2021171949A1) covers the data and applications of the findings described in this manuscript.

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer files 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 General features of sEV secretion among the five cell types studied.](#)

a) Number of vesicles over 48 h released by each cell type normalized by the number of cells in the tissue-culture plate (n=4). *P≤0.05 (indicated cell

type versus all other cell types), § $P \leq 0.05$ (indicated cell type versus 3T3-L1, C2C12 and SVEC) (Kruskal-Wallis followed by Mann-Whitney U test). **b)** Average vesicle size of the sEV as determined by Nanoparticle tracking analysis (NTA) for each cell type ($n=4$). **c)** Average size distribution and number of vesicles released per cell for each of the five cell types. **d)** Immunoblotting for the indicated sEV (ALIX, TSG101 and CD9) and cellular (GM130, CANX) markers in sEV and cell lysates from AML12 hepatocytes and BAT brown adipocytes. **e)** Electron micrograph showing CD63 gold immunostaining of sEV isolated from C2C12 cells. **f)** RNA yield obtained from sEV isolated from each cell type and normalized by the number of cells in the tissue-culture plate ($n=3$). * $P \leq 0.05$ (indicated cell type versus all other cell types); § $P \leq 0.05$ (indicated cell type versus 3T3-L1 and BAT) (Kruskal-Wallis followed by Mann-Whitney U test). **g)** Principal component analysis showing cellular miRNA profiles for each cell type. **h)** Heatmap showing the top 10 representative cellular miRNAs of each cell type. **i)** Heatmap showing the top 10 representative sEV miRNAs of each cell type. **j)** Comparative miRNA profile between cell-derived sEV and non-conditioned medium (NCM). Same volume of NCM as in cell-conditioned medium was processed for sEV isolation by differential ultracentrifugation. RNA was isolated and a miRNA profiling was performed for NCM. The miRNA expressions for the 13 miRNAs found sEV-enriched in all 5 cell types were compared to the NCM average Ct by $\Delta\Delta ct$ method and represented as fold change. Each dot is the relative average value of each of the five cell types. Data are expressed as mean \pm SEM.

Extended Data Fig. 2 Cell-type specific sEV enrichment of miRNAs and sEV versus cell housekeeping miRNAs. Effect of percentage of CG and Gibbs free energy (ΔG) in miRNA sorting or cellular retention.

a) Normalized gene expression of representative miRNAs showing cell-type specific sEV enrichment: miR-696 (BAT), miR-770-5p (C2C12), miR-1927 (3T3-L1), miR-1931 (AML12) and miR-718 (SVEC). Average Ct from the whole miRNA profile was used for normalization for each sample ($n=3-4$). * $P \leq 0.05$ (Limma t-test). **b)** Normalized gene expression of two

representative miRNAs (miR-138-5p and miR-501-5p) showing similar sEV and cellular expression for each cell type. Average Ct from the whole miRNA profile was used for normalization for each sample (n=3-4). **c**) Percentage of CG content in the sequence of the miRNAs: those sorted into the sEV in all cells are shown in red (n=13); those sorted into sEV in 3 or 4 of the five cell types are shown in green (n=90); those not enriched in either sEV or cells are shown in black (n=109); those retained in 3 or 4 cell types are shown in pink (n=97); and those retained in the cell bodies of all cell types are shown in blue (n=43). **d**) The calculated Gibbs free energy (ΔG) for each of the miRNA in each of the five groups described in panel c. For **c**, **d** lines indicate mean value, ** $P \leq 0.01$, *** $P \leq 0.001$ between the indicated group and the other four groups (ANOVA followed by Bonferroni post-hoc test in **c**, **d**). Data are expressed as mean \pm SEM.

Extended Data Fig. 3 Comparison of all identified Extended and Core EXOmotifs (a) and CELLmotifs (b) among the five different cell types.

a) Fold enrichment (left half) and abundance (right half, as percentage of sEV-enriched miRNAs) containing the sEV-associated miRNA motifs. The first column indicates the cell type where these motifs were originally identified (see Main Fig. 2a) and here shown in its predominant form in the second column (Extended motifs) and fifth column (Core motifs). The fold enrichment is shown in a red (high sEV enrichment)-white (neutral)-blue (cell enrichment) color gradient for indicated cell types displayed below. The abundance is shown in a red (high)-white (low) color gradient. The rectangles highlight the enrichment and presence of the motifs in the cell types where they were originally identified. **b)** Fold enrichment (left half) and abundance (right half, as percentage of cell-enriched miRNAs) containing the Cell-associated miRNA motifs. As in a), the first column indicates the cell type where these motifs were originally identified (see Main Fig. 2b) and here shown in its predominant form in the second column (Extended motifs) and fifth column (Core motifs). The fold enrichment is shown in a blue (high cell enrichment)-white (neutral)-blue (red sEV enrichment) color gradient for indicated cell types displayed below. The abundance is shown in a blue (high)-white (low) color gradient.

The rectangles highlight the enrichment and presence of the motifs in the cell types where they were originally identified.

Extended Data Fig. 4 Location of EXO and CELLmotifs and comparison between small miRNAs (smRNAs) and qPCR-based miRNA profiling.

a, b) Percentage of miRNAs showing indicated EXOmotifs (a) or CELLmotifs (b) either in the 5' half (nucleotides 1-9, light yellow bars) or 3' half (from nucleotide 10 to the 3' end, orange bars) of the miRNA sequence. **c)** sEV and cell lysates from differentiated 3T3-L1 white adipocytes and AML12 hepatocytes were subjected to smRNAs (smRNAs) or qPCR-based profiling (n=4 for each cell type and compartment). Detected miRNAs by each method in each cell type correspond to the sum of black (non-selectively distributed miRNAs) and blue bars (selectively distributed miRNAs either in sEV-enriched or cell-enriched). The percentages in the blue bars refer to the ratio between the number of selectively distributed miRNA and the total number of detected miRNAs for each method and cell type. **d)** Venn diagrams indicating the number of miRNAs with a selective distribution in sEV or cells detected by smRNAs (blue circles) and qPCR (green circles) in 3T3-L1 (above) or AML12 (below). The total number of miRNAs detected simultaneously by these two methods was 180. **e)** Table depicting the top EXOmotif found by HOMER software in sEV-enriched miRNAs from 3T3-L1 and AML12 detected by smRNAs. Fold enrichment refers to the ratio between presence in the sEV-enriched miRNAs and presence in the rest of miRNAs (background).

Extended Data Fig. 5 Comparison of miRNA profiling of AML12 and primary hepatocytes.

Primary hepatocytes were isolated from C57Bl/6J wild-type mice (n=4) and cultured for 48 h in exosome-free medium to collect sEV and cell lysates, which were later subjected to RNA isolation and miRNA profiling for comparison to AML12 hepatocytes. **a)** Number of miRNAs showing selective cellular retention, non-selective distribution or selective sEV sorting in AML12 and primary hepatocytes. The selective distribution

column is the sum of sEV- and cell-enriched miRNAs divided by the total number of miRNAs detected as percentage. FDR < 0.1. **b**) Venn-diagram showing the number of sEV- (red) and cell-enriched miRNAs (blue) in AML12 and primary hepatocytes and the overlap between them. **c, d)** Motifs associated to sEV (**c**) and cell (**d**) enrichment in primary hepatocytes. The table shows the significance of the enrichment (P-value), false discovery rate (FDR), the percentage of miRNAs significantly enriched in sEV (in **c**) or cell (in **d**) that contain the motif, the percentage of miRNAs not enriched in the background miRNAs containing the motif and the fold-enrichment as the ratio between the previous two columns.

Extended Data Fig. 6 Isolation of sEV and NV using an additional step of size exclusion chromatography.

a) Diagram of the isolation method used to obtain cellular, sEV-p100 pellet, sEV-SEC and NV-SEC samples using two rounds of ultracentrifugation followed by size exclusion chromatography (SEC). (n=4). **b)** NTA analysis for EV concentration (top graph) of the 30 fractions obtained from SEC. These fractions were pooled in pairs in some cases and concentrated using Amicon centrifugal 3 KDa filter prior assessing protein concentration (bottom graph). **c)** Immunoblot for classical exosomal markers CD63 and CD9 for the concentrated fractions shown in b, bottom graph. **d)** PCA plot for the miRNA profile from the cells, sEV-p100, sEV-SEC and NV-SEC samples. **e)** Heatmap of the top differentially expressed miRNAs among sEV-p100, sEV-SEC and NV-SEC. High expression is shown in red and low expression in blue. **f)** Pearson correlation between averaged normalized miRNA expression levels in sEV-p100 and sEV-SEC. Expression levels were first normalized to average ct of each sample. **g)** Motifs found overrepresented in sEV-SEC enriched-miRNAs compared to cellular-enriched miRNAs.

Extended Data Fig. 7 Additional information about mutations in miR-431-5p, miR-140-3p and miR-677-5p.

a) Table depicting the name and sequence of the CELLmotif AGAAC incorporated in miR-431-5p. Bold underlying text in the sequence indicated

changed nucleotides in the guide strand of the miRNA used to introduce the CELLmotif AGAAC. Nucleotides in the passenger strand were also modified to maintain miRNA structure. **b)** Predicted structure for the hairpin miRNA for the constructs shown in a. Red means high probability of pairing, while blue indicates low probability calculated by RNAfold WebServer software. Arrow indicates the location of the mutated nucleotides. **c)** Normalized gene expression for miR-431-5p wild-type in cells and sEV for each cell type. Average Ct from the whole miRNA profile was used for normalization for each sample. **d)** Table depicting the name and sequence of the wild-type miR-140-3p and mutated version without CELLmotif AGAAC. Bold underlying text in the sequence indicated changed nucleotides in the guide strand of the miRNA used to remove the CELLmotif AGAAC. Nucleotides in the passenger strand were also modified to maintain miRNA structure. **e)** Predicted structure for the hairpin miRNA for the constructs shown in d. Red means high probability of pairing, while blue indicates low probability. Arrow indicates the location of the mutated nucleotides. **f)** Normalized gene expression for miR-140-3p wild-type in cells and sEV for each cell type. Average Ct from the whole miRNA profile was used for normalization for each sample. **g)** Table showing the name and sequence of the wild-type miR-677-5p and the version in which CELLmotifs were mutated. Bold underlined text in the sequence indicated mutated nucleotides in the guide strand of the miRNA to remove CELLmotifs CAGU and AUU[A/G]. Nucleotides in the passenger strand were also modified accordingly to maintain miRNA structure. **h)** Predicted structure for the hairpin miRNA for the constructs shown in g. Red means high probability of pairing while blue indicates low probability. Arrows indicate the location of the mutated nucleotides. **i)** Normalized gene expression for miR-677-5p wild-type in cells and exosomes/sEV for each cell type. Average Ct from the whole miRNA profile was used for normalization for each sample. Data are expressed as mean \pm SEM. n=3-4.
*P \leq 0.05 (Limma t-test).

Extended Data Fig. 8 Additional information about mutations in miR-34c.

a) Table depicting the name and sequence of the different EXOmotifs introduced in miR-34c-5p. Bold underlying text in the sequence indicates

changed nucleotides in the guide strand of the miRNA. Nucleotides in the passenger strand were also modified accordingly to maintain miRNA structure. **b)** Predicted structure for the hairpin miRNA for each of the constructs shown in a. Red means high probability of pairing while blue indicates low probability calculated by RNAfold WebServer software. Arrows indicate the location of the mutated nucleotides. **c)** Normalized expression for miR-34c-5p wild-type in cells and sEV for each cell type. Average Ct from the whole miRNA profile was used for normalization for each sample. * $P \leq 0.05$ (Limma t-test). **d)** To visualize the changes in the sEV and cell content of each miRNA construct before sEV enrichment calculation, normalized expression of miR-34c wild-type (WT) or its EXOmotif-containing versions miR-34-UGUGU, miR-34-CAUG and miR-34-CGGGAG are displayed for the cells overexpressing each of the miR-34c versions. miR-138-5p was used to normalize expression in sEV versus cells as we previously showed that the levels of this miRNA are equivalent in both compartments.* $P \leq 0.05$ (Mann-Whitney U test comparing sEV and cell expression for the same miRNA version, i.e. red versus blue bars). **e)** Absolute copy number of the miRNAs displayed in the x-axis overexpressed in brown adipocytes were quantified in sEV isolated by ultracentrifugation followed by size exclusion chromatography and normalized by the miRNA copy number in all producing cells for each sample. **f)** sEV enrichment calculated as the ratio of sEV expression divided by cellular expression for each of the constructs expressed in and secreted from AML12 hepatocytes. **g)** sEV enrichment calculated as the ratio of sEV expression divided by cellular expression for each of the constructs expressed in and secreted from SVEC endothelial cells. In both panels **f**, **g**, the dashed line separates preferential sEV enrichment (above line) versus preferential cellular enrichment (below line). Expression was normalized to the expression of miR-501-5p, which is to be equally abundant in sEV and cells for each cell type. Data are expressed as mean \pm SEM. n=3, * $P \leq 0.05$, Kruskal-Wallis followed by Mann-Whitney U tests.

Extended Data Fig. 9 Additional information about mutations in miR-26a.

a) Table depicting the name and sequence of the different EXOmotifs introduced in miR-26a-5p. Bold underlying text in the sequence indicates

changed nucleotide/s in the guide strand of the miRNA. Nucleotides in the passenger strand were also modified accordingly to maintain miRNA structure. **b)** Predicted structure for the hairpin miRNA for each of the constructs shown in a. Red means high probability of pairing while blue indicates low probability calculated by RNAfold WebServer software. Arrows indicate the location of the mutated nucleotide/s. **c)** Normalized expression for miR-26a-5p wild-type in cells and sEV for each cell type. Average Ct from the whole miRNA profile was used for normalization for each sample. *P≤0.05 (Limma t-test). **d)** sEV enrichment calculated as the ratio of sEV expression divided by cellular expression for each of the constructs expressed in and secreted from SVEC endothelial cells. The dashed line separates preferential sEV enrichment (above line) versus preferential cellular enrichment (below line). Expression was normalized to the expression of miR-501-5p. *P≤0.05 (Kruskal-Wallis followed by Mann-Whitney U test), n=3-4.

Extended Data Fig. 10 Further information miRNA pulldown and direct transfection of AML12 cells.

a) Analyses for Molecular Function (left) and Cellular Component Gene Ontology (right) of the 67 proteins identified in the proteomic study. **b)** Table showing average values for relative binding enrichment of the proteins listed in the first column to the miRNA constructs shown in the top row. The columns 2-5 refer to miR-34c and its CGGGAG-containing version, while columns 6-8 refer to miR-26a and its CGGGAG-containing version. In both cases, binding to wild-type miRNAs was set as 1, and the binding of the other miRNA constructs (scramble and CGGGAG-containing version) was normalized respect to that. Only those proteins showing a log₂ Fold Enrichment of CGGGAG-containing miRNA version versus wild-type miRNA >3 (>8 fold) were included. **c)** Brown adipocytes overexpressing wild type miR-34c (OE-miR-34c-5p-WT, grey bars) or CGGGAG-containing miR-34c (OE-miR-34c-5p-CGGGAG, red bars) were transfected with either control siRNA, Alyref siRNA or Fus siRNA (as indicated in x-axes) and analyzed by qPCR for knockdown efficiency for Alyref (left graph) or Fus (right graph), or by immunoblotting (right). *P≤0.05 (Kruskal-Wallis followed by Mann-Whitney U test), n=3. **d)** EXOmotif-containing miR-34c versions have the same efficiency in

reducing target gene expression than wild-type miR-34c. AML12 hepatocytes were directly transfected with mimic miR-34c wild-type or its mutant versions miR-34-UGUGU, miR-34-CAUG and miR-34-CGGGAG or non-targeting miRNA (control) for 24 h and expression of predicted and experimentally-validated miR-34c target genes were analyzed. TATA-box binding protein (*Tbp*) was used as housekeeping gene. n=6. *P≤0.05, ***P≤0.001 (ANOVA followed by Bonferroni post-hoc test). Data are expressed as mean ± SEM.

Supplementary information

Supplementary Fig. 1

This file contains uncropped gels for the images shown in Extended Data Figs. 1, 6 and 10.

Reporting Summary

Peer Review Information

Supplementary Information

This file contains the legends for Supplementary Tables 1–11.

Supplementary Table 1

Negative ΔC_t value for each miRNA normalized by average C_t in sEVs/exosomes and cell bodies obtained from miRNA profiling.

Supplementary Table 2

Average normalized raw expression and statistics of each miRNA in cell bodies for each cell type.

Supplementary Table 3

Significance table of cell-type-specific miRNAs for each cell type.

Supplementary Table 4

Average normalized raw expression and statistics of each miRNA in exosomes/sEVs for each cell type.

Supplementary Table 5

Significance table of exosome/sEV-specific miRNAs for each cell type.

Supplementary Table 6

sEV enrichment miRNAs.

Supplementary Table 7

Primary hepatocyte miRNA profile.

Supplementary Table 8

Ultracentrifugation + SEC miRNA profile.

Supplementary Table 9

List of proteins from the miRNA pulldown experiment and their statistics.

Supplementary Table 10

Expression (as counts per million, CPM) of each detected miRNA in the smRNA-seq experiment.

Supplementary Table 11

Pre-miRNA sequences showing all mutations in both guide and passenger strands.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Garcia-Martin, R., Wang, G., Brandão, B.B. *et al.* MicroRNA sequence codes for small extracellular vesicle release and cellular retention. *Nature* **601**, 446–451 (2022). <https://doi.org/10.1038/s41586-021-04234-3>

- Received: 25 December 2019
- Accepted: 12 November 2021
- Published: 22 December 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04234-3>

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

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

- Article
- [Published: 15 December 2021](#)

Synthon-based ligand discovery in virtual libraries of over 11 billion compounds

- [Arman A. Sadybekov](#) ORCID: [orcid.org/0000-0002-0573-4738^{1,2 na1}](https://orcid.org/0000-0002-0573-4738),
- [Anastasiia V. Sadybekov](#) ORCID: [orcid.org/0000-0003-3925-983X^{1,2 na1}](https://orcid.org/0000-0003-3925-983X),
- [Yongfeng Liu](#) ORCID: [orcid.org/0000-0002-8649-7892^{3,6 na1}](https://orcid.org/0000-0002-8649-7892),
- [Christos Iliopoulos-Tsoutsouvas](#)^{5 na1},
- [Xi-Ping Huang](#)^{3,6},
- [Julie Pickett](#) ORCID: [orcid.org/0000-0002-9535-8528^{3,6}](https://orcid.org/0000-0002-9535-8528),
- [Blake Houser](#)²,
- [Nilkanth Patel](#)¹,
- [Ngan K. Tran](#)⁵,
- [Fei Tong](#)⁵,
- [Nikolai Zvonok](#)⁵,
- [Manish K. Jain](#) ORCID: [orcid.org/0000-0001-8173-3326³](https://orcid.org/0000-0001-8173-3326),
- [Olena Savych](#)⁷,
- [Dmytro S. Radchenko](#) ORCID: [orcid.org/0000-0001-5444-7754^{7,8}](https://orcid.org/0000-0001-5444-7754),
- [Spyros P. Nikas](#)⁵,
- [Nicos A. Petasis](#) ORCID: [orcid.org/0000-0002-8864-4446²](https://orcid.org/0000-0002-8864-4446),
- [Yuri S. Moroz](#) ORCID: [orcid.org/0000-0001-6073-002X^{8,9}](https://orcid.org/0000-0001-6073-002X),
- [Bryan L. Roth](#) ORCID: [orcid.org/0000-0002-0561-6520^{3,4,6}](https://orcid.org/0000-0002-0561-6520),
- [Alexandros Makriyannis](#) ORCID: [orcid.org/0000-0003-3272-3687^{5,10}](https://orcid.org/0000-0003-3272-3687) &
- [Vsevolod Katritch](#) ORCID: [orcid.org/0000-0003-3883-4505^{1,2}](https://orcid.org/0000-0003-3883-4505)

[Nature](#) volume **601**, pages 452–459 (2022)

- 13k Accesses
- 4 Citations
- 166 Altmetric
- [Metrics details](#)

Subjects

- [Cheminformatics](#)
- [Diversity-oriented synthesis](#)
- [Molecular modelling](#)
- [Molecular neuroscience](#)
- [Virtual screening](#)

Abstract

Structure-based virtual ligand screening is emerging as a key paradigm for early drug discovery owing to the availability of high-resolution target structures^{1,2,3,4} and ultra-large libraries of virtual compounds^{5,6}. However, to keep pace with the rapid growth of virtual libraries, such as readily available for synthesis (REAL) combinatorial libraries⁷, new approaches to compound screening are needed^{8,9}. Here we introduce a modular synthon-based approach—V-SYNTHES—to perform hierarchical structure-based screening of a REAL Space library of more than 11 billion compounds. V-SYNTHES first identifies the best scaffold–synthon combinations as seeds suitable for further growth, and then iteratively elaborates these seeds to select complete molecules with the best docking scores. This hierarchical combinatorial approach enables the rapid detection of the best-scoring compounds in the gigascale chemical space while performing docking of only a small fraction (<0.1%) of the library compounds. Chemical synthesis and experimental testing of novel cannabinoid antagonists predicted by V-SYNTHES demonstrated a 33% hit rate, including 14 submicromolar ligands, substantially improving over a standard virtual screening of the

Enamine REAL diversity subset, which required approximately 100 times more computational resources. Synthesis of selected analogues of the best hits further improved potencies and affinities (best inhibitory constant (K_i) = 0.9 nM) and CB₂/CB₁ selectivity (50–200-fold). V-SYNTHES was also tested on a kinase target, ROCK1, further supporting its use for lead discovery. The approach is easily scalable for the rapid growth of combinatorial libraries and potentially adaptable to any docking algorithm.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: V-SYNTHES approach to modular screening of Enamine REAL Space.

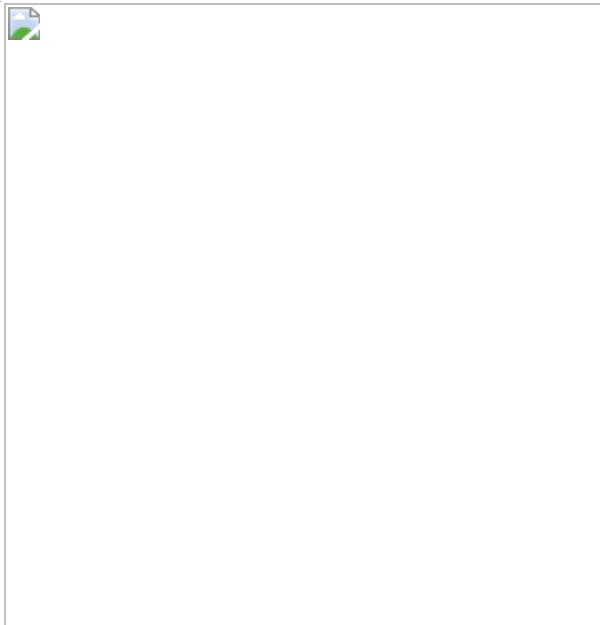


Fig. 2: Assessment of VLS computational performance for V-SYNTHES and standard VLS.

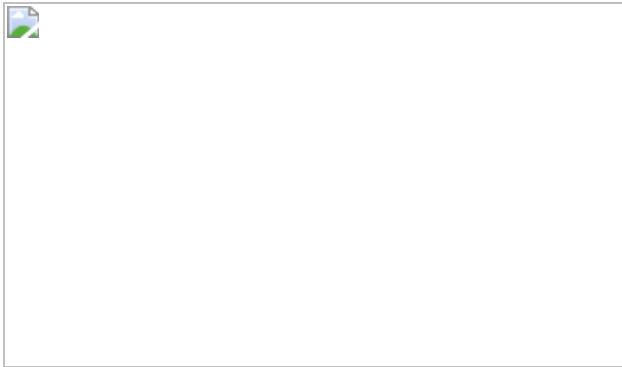


Fig. 3: The top five CB₂ hits identified by V-SYNTHES.

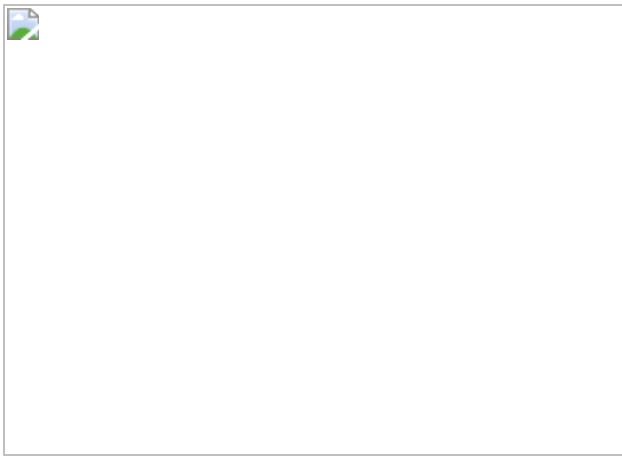
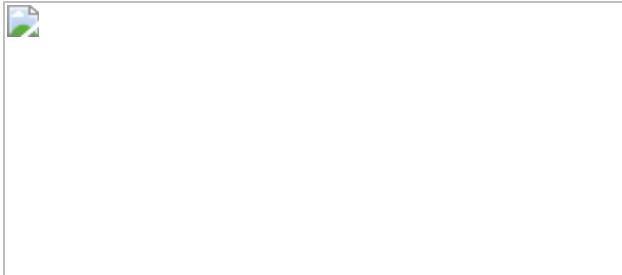


Fig. 4: Selection and characterization of the best analogue series for CB₂ hits from V-SYNTHES screening.



Data availability

Chemical structures, synthetic methods, detailed results of biochemical characterization are presented in this paper and its [Supplementary Information](#).

Code availability

V-SYNTHES scripts and example files have been deposited at GitHub (<https://github.com/katritchlab/V-SYNTHES>).

References

1. 1.
Shoichet, B. K. & Kobilka, B. K. Structure-based drug screening for G-protein-coupled receptors. *Trends Pharmacol. Sci.* **33**, 268–272 (2012).
2. 2.
Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).
3. 3.
Renaud, J.-P. et al. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat. Rev. Drug Discov.* **17**, 471–492 (2018).
4. 4.
Congreve, M., de Graaf, C., Swain, N. A. & Tate, C. G. Impact of GPCR structures on drug discovery. *Cell* **181**, 81–91 (2020).
5. 5.
Stein, R. M. et al. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 609–614 (2020).
6. 6.
Lyu, J. et al. Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).
7. 7.

Grygorenko, O. O. et al. Generating multibillion chemical space of readily accessible screening compounds. *iScience* **23**, 101681 (2020).

8. 8.

Gorgulla, C. et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663–668 (2020).

9. 9.

Graff, D. E., Shakhnovich, E. I. & Coley, C. W. Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem. Sci.* **12**, 7866–7881 (2021).

10. 10.

Engels, M. F. & Venkatarangan, P. Smart screening: approaches to efficient HTS. *Curr. Opin. Drug Discov. Dev.* **4**, 275–283 (2001).

11. 11.

Villoutreix, B. O., Eudes, R. & Miteva, M. A. Structure-based virtual ligand screening: recent success stories. *Comb. Chem. High Throughput Screen.* **12**, 1000–1016 (2009).

12. 12.

Abagyan, R. & Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **5**, 375–382 (2001).

13. 13.

Irwin, J. J. & Shoichet, B. K. Docking screens for novel ligands conferring new biology. *J. Med. Chem.* **59**, 4103–4120 (2016).

14. 14.

Ertl, P. Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties,

and automatic identification of drug-like bioisosteric groups. *J. Chem. Inf. Comput. Sci.* **43**, 374–380 (2003).

15. 15.

Bohacek, R. S., McMurtin, C. & Guida, W. C. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* **16**, 3–50 (1996).

16. 16.

REAL Space (Enamine, 2020); <https://enamine.net/library-synthesis/real-compounds/real-space-navigator>

17. 17.

Guzmán, M. Cannabinoids: potential anticancer agents. *Nat. Rev. Cancer* **3**, 745–755 (2003).

18. 18.

Contino, M., Capparelli, E., Colabufo, N. A. & Bush, A. I. Editorial: the CB2 cannabinoid system: a new strategy in neurodegenerative disorder and neuroinflammation. *Front. Neurosci.* **11**, 196 (2017).

19. 19.

Lunn, C. A. et al. Biology and therapeutic potential of cannabinoid CB2 receptor inverse agonists. *Br. J. Pharmacol.* **153**, 226–239 (2008).

20. 20.

Corey, E. J. General methods for the construction of complex molecules. *Pure Appl. Chem.* **14**, 19–38 (1967).

21. 21.

Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries

and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740 (2010).

22. 22.

Li, X. et al. Crystal structure of the human cannabinoid receptor CB2. *Cell* **176**, 459–467 (2019).

23. 23.

Kroeze, W. K. et al. PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biol.* **22**, 362–369 (2015).

24. 24.

Gaulton, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).

25. 25.

Xing, C. et al. Cryo-EM structure of the human cannabinoid receptor CB2-Gi signaling complex. *Cell* **180**, 645–654 (2020).

26. 26.

Wei, L., Surma, M., Shi, S., Lambert-Cheatham, N. & Shi, J. Novel insights into the roles of rho kinase in cancer. *Arch. Immunol. Ther. Exp.* **64**, 259–278 (2016).

27. 27.

Chin, V. T. et al. Rho-associated kinase signalling and the cancer microenvironment: novel biological implications and therapeutic opportunities. *Expert Rev. Mol. Med.* **17**, e17 (2015).

28. 28.

Baker, M. Fragment-based lead discovery grows up. *Nat. Rev. Drug Discov.* **12**, 5–7 (2013).

29. 29.

Schulz, M. N. & Hubbard, R. E. Recent progress in fragment-based lead discovery. *Curr. Opin. Pharmacol.* **9**, 615–621 (2009).

30. 30.

Davis, B. J. & Hubbard, R. E. in *Structural Biology in Drug Discovery* 79–98 (2020).

31. 31.

Zheng, Z. et al. Structure-based discovery of new antagonist and biased agonist chemotypes for the kappa opioid receptor. *J. Med. Chem.* **60**, 3070–3081 (2017).

32. 32.

de Graaf, C. et al. Crystal structure-based virtual screening for fragment-like ligands of the human histamine H₁ receptor. *J. Med. Chem.* **54**, 8195–8206 (2011).

33. 33.

Katritch, V. et al. Structure-based discovery of novel chemotypes for adenosine A_{2A} receptor antagonists. *J. Med. Chem.* **53**, 1799–1809 (2010).

34. 34.

Chen, Y. & Shoichet, B. K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **5**, 358–364 (2009).

35. 35.

Abagyan, R. A., Orry, A., Raush, E., Budagyan, L. & Totrov, M. *ICM User's Guide and Reference Manual* v.3.9 (MolSoft, 2021).

36. 36.

Bogolubsky, A. V. et al. A one-pot parallel reductive amination of aldehydes with heteroaromatic amines. *ACS Comb. Sci.* **16**, 375–380 (2014).

37. 37.

Savych, O. et al. One-pot parallel synthesis of 5-(dialkylamino)tetrazoles. *ACS Comb. Sci.* **21**, 635–642 (2019).

38. 38.

Katritch, V., Rueda, M. & Abagyan, R. Ligand-guided receptor optimization. *Methods Mol. Biol.* **857**, 189–205 (2012).

39. 39.

Gatica, E. A. & Cavasotto, C. N. Ligand and decoy sets for docking to G protein-coupled receptors. *J. Chem. Inf. Model.* **52**, 1–6 (2012).

40. 40.

Bottegoni, G., Kufareva, I., Totrov, M. & Abagyan, R. Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. *J. Med. Chem.* **52**, 397–406 (2009).

41. 41.

Real Compound Libraries (Enamine, 2020);
<https://enamine.net/library-synthesis/real-compounds/real-compound-libraries>

42. 42.

Nikas, S. P. et al. Probing the carboxyester side chain in controlled deactivation ($-$)- Δ^8 -tetrahydrocannabinols. *J. Med. Chem.* **58**, 665–681 (2015).

43. 43.

Nikas, S. P. et al. Novel 1',1'-chain substituted hexahydrocannabinols: 9 β -hydroxy-3-(1-hexyl-cyclobut-1-yl)-hexahydrocannabinol (AM2389) a highly potent cannabinoid receptor 1 (CB1) agonist. *J. Med. Chem.* **53**, 6996–7010 (2010).

44. 44.

Jacobs, M. et al. The structure of dimeric ROCK I reveals the mechanism for ligand selectivity. *J. Biol. Chem.* **281**, 260–268 (2006).

45. 45.

Anastassiadis, T., Deacon, S. W., Devarajan, K., Ma, H. & Peterson, J. R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat. Biotechnol.* **29**, 1039–1045 (2011).

Acknowledgements

We thank the staff at the USC Center for Advanced Research Computing, and the Google Cloud Platform for Higher Education and Research for providing computational resources. The study was funded by National Institute on Drug Abuse grants R01DA041435 and R01DA045020 (to V.K. and A.M.), National Institute of Mental Health Grant R01MH112205 and Psychoactive Drug Screening Program (to B.L.R.) and the Michael Hooker Distinguished Professorship (to B.L.R.). B.H. was supported by NIGMS T32-GM118289.

Author information

Author notes

1. These authors contributed equally: Arman A. Sadybekov, Anastasia V. Sadybekov, Yongfeng Liu, Christos Iliopoulos-Tsoutsouvas

Affiliations

1. Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA

Arman A. Sadybekov, Anastasia V. Sadybekov, Nilkanth Patel & Vsevolod Katritch

2. Department of Chemistry, Bridge Institute, USC Michelson Center for Convergent Biosciences, University of Southern California, Los Angeles, CA, USA

Arman A. Sadybekov, Anastasia V. Sadybekov, Blake Houser, Nicos A. Petasis & Vsevolod Katritch

3. Department of Pharmacology, School of Medicine, University of North Carolina, Chapel Hill, NC, USA

Yongfeng Liu, Xi-Ping Huang, Julie Pickett, Manish K. Jain & Bryan L. Roth

4. Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

Bryan L. Roth

5. Center for Drug Discovery, Department of Pharmaceutical Sciences, Northeastern University, Boston, MA, USA

Christos Iliopoulos-Tsoutsouvas, Ngan K. Tran, Fei Tong, Nikolai Zvonok, Spyros P. Nikas & Alexandros Makriyannis

6. Psychoactive Drug Screening Program, National Institute of Mental Health, School of Medicine, University of North Carolina, Chapel Hill,

NC, USA

Yongfeng Liu, Xi-Ping Huang, Julie Pickett & Bryan L. Roth

7. Enamine Ltd, Kyiv, Ukraine

Olena Savych & Dmytro S. Radchenko

8. Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Dmytro S. Radchenko & Yurii S. Moroz

9. Chemspace LLC, Kyiv, Ukraine

Yurii S. Moroz

10. Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA

Alexandros Makriyannis

Contributions

A.A.S. and A.V.S. developed V-SYNTHES algorithms, performed calculations and wrote the first draft of the manuscript. B.H. and N.A.P. performed calculations and compound selection for ROCK1. Y.L., M.K.J., J.P. and X.-P.H. performed functional and selectivity assays. C.I.-T., N.K.T., F.T., N.Z. and S.P.N. performed binding assays. N.P. performed full VLS on Google Cloud. O.S., D.S.R. and Y.S.M. developed the REAL Space library and performed compound synthesis. B.L.R. supervised the functional and selectivity assays. A.M. supervised binding assays for CB₁ and CB₂. V.K. conceived the study and supervised all of its computational aspects. All of the authors contributed to writing and editing the manuscript.

Corresponding authors

Correspondence to [Bryan L. Roth](#), [Alexandros Makriyannis](#) or [Vsevolod Katritch](#).

Ethics declarations

Competing interests

A.A.S. and V.K. filed a provisional patent on V-SYNTHES method (application no. 63159888, University of Southern California).

Additional information

Peer review information *Nature* thanks Charlotte Dean and Amy Newman for their contribution to the peer review of this work.

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

Extended data figures and tables

[Extended Data Fig. 1 Evaluation of SYNTHEs performance on CB₂ receptor with only docking score \(without considering docking pose of MEL candidates in the binding pocket\).](#)

(a) The number of hits at each score threshold from V-SYNTHEs and standard VLS (b) Enrichment in V-SYNTHEs vs. Standard VLS at different score thresholds, with the red x-mark showing threshold that yields 100 V-SYNTHEs hits in the two-component library.

[Extended Data Fig. 2 Binding pocket of CB₂ with selected dead-end atoms.](#)

a) 3D illustration of a MEL compound binding pose (carbon atoms colored cyan) with a “non-productive” pose. (b-d) 2D schematics showing other possible non-productive cases, including dead-end subpockets. Dead-end water-colored red, pseudoatoms colored magenta.

Extended Data Fig. 3 Details of practical application V-SYNTHES algorithms to CB receptors screening.

a, b, Two-component (**a**) and three-component (**b**) reaction cases.

Extended Data Fig. 4 Concentration-response curves for V-SYNTHES hits in functional assays at CB₁ and CB₂ receptors (except those shown in main text Figure 3).

β -arrestin recruitment Tango assays were performed to assess antagonist activity of the compounds in (**a,b**) CB₁ and (**c,d**) CB₂ receptors. The compounds rimonabant or SR144528 served as positive controls. The assays were carried out in the presence of 100 nM (EC₈₀) of the dual CB₁/CB₂ CP55,940 agonist. The data points are presented as mean \pm SEM with n = 3 independent experiments, each one carried out in triplicate.

Extended Data Fig. 5 Competition binding curves for the best CB₂ hit compounds from V-SYNTHES.

Radioligand binding assays were used to assess the binding affinities in rCB1 (**a**) and hCB2 (**b**). [³H]CP-55,940 was used as the radioligand. The data were presented as mean \pm SEM with n = 3 independent experiments, each one carried out in triplicate.

Extended Data Fig. 6 Assessment of off-target selectivity for the best V-SYNTHES CB₂ hits.

(**a-c**) Screening of compounds 673, 610 and 523 at 10 μ M concentrations in GPCRome-Tango assays for >300 receptors. Dopamine D₂ (DRD2) and 100 nM Quinpirole served as an assay control. The data are presented as mean \pm SEM (n = 4) and the values of fold of basal > 3 are marked as significant hits. (**d-o**) Follow-up dose-response curves for targets with >3 fold increased activity. Known agonists or antagonist that showed activity

served as positive controls. The data points are presented as mean ± SEM with n = 3 independent experiments, each assay carried out in triplicate.

Extended Data Fig. 7 Identification and characterization of CB₁ and CB₂ hits from standard VLS of 115M Enamine REAL compounds.

(a) Chemical structures of the hits from the standard VLS. (b-c). Concentration-response curves of the best hits in β-arrestin recruitment Tango assays for antagonist activity at CB₁ (b) and CB₂ (c) receptors. The compounds rimonabant or SR144528 served as positive controls. The assays were carried out in the presence of 100 nM (EC₈₀) of the dual CB₁/CB₂ CP55,940 agonist. The data points are presented as mean ± SEM with n = 3 independent experiments, each one carried out in triplicate. (d) Functional potencies and binding affinities of the hit compounds from standard VLS. The 95% Confidence Intervals (CI) were calculated from n = 3 independent assays, with 16 dose-response points for functional K_i values and 8 dose-response points for affinity K_i values, except for values marked with *, roughly estimated from three-point assays.

Extended Data Fig. 8 Competition binding curves for the best CB₂ hit compounds from standard VLS.

Radioligand binding assays were used to assess the binding affinities in hCB2. [³H]CP-55,940 was used as the radioligand. The data were presented as mean ± SEM with n = 3 independent experiments, each one carried out in triplicate.

Extended Data Fig. 9 Chemical structures for series of the SAR-by-catalog analogues of antagonists, discovered by V-SYNTHES.

Shown are 60 analogues of 523 (a), 610 (b), and 673 (c) with inhibitory activity >40% in the single point functional assays. All 104 analogues tested are shown in Supplementary Information Table S3.

Extended Data Fig. 10 Functional potency and binding affinity assessment of the SAR-by-catalog analogues of the antagonist 523, discovered by V-SYNTHES.

Table compounds with CB₂ potency better than 500 nM are shown, antagonists with affinities better than 10 nM highlighted in bold, >50-fold selective by italic. Functional K_i values and 95% Confidence Intervals were calculated from n = 4 independent assays with 16 dose-response points. Affinity Ki values and 95% Confidence Intervals were calculated from n = 3 independent assays with 8 dose-response points.

Extended Data Fig. 11 Concentration-response curves for series of the SAR-by-catalog analogues of 523, 610 and 673 antagonists, discovered by V-SYNTHES.

The β-arrestin recruitment Tango assays were performed to assess the antagonist activity of the best hits at CB₁ (**a-i**), and CB₂ (**j-o**) receptors. Note that the six best analogues of **523** shown in Fig. [4](#) are excluded here. The compounds rimonabant and SR144528 served as positive controls. The assays were carried out in the presence of 100 nM (EC₈₀) of the CP55,940 agonist. The data were presented as mean ± SEM with n = 3 independent experiments, each run carried out in triplicate.

Extended Data Fig. 12 Assessment of off-target selectivity for the best SAR-by-catalog compounds 733 and 747.

(**a-b**) Screening of compounds 733 and 747 in GPCRome-Tango assay for >300 receptors at 10 μM concentrations. Dopamine D₂ (DRD2) and 100 nM Quinpirole served as an assay control. The data are presented as mean ± SEM (n = 4) and the values of fold of basal > 3 marked as significant hits.
(**c-d**) Follow-up dose-response curves for targets with >3 fold increased activity. Known agonists that showed activity served as positive controls. The data were presented as mean ± SEM with n = 3 independent experiments, each run carried out in triplicate.

Extended Data Fig. 13 Application of V-SYNTHES to the discovery of ROCK1 inhibitors.

(a,b) Computational assessment of V-SYNTHES performance vs standard VLS. **(a)** The number of candidate hits at each score threshold from V-SYNTHES and standard VLS. **(b)** Enrichment in V-SYNTHES vs. standard VLS at different score thresholds, with the red x-mark showing threshold that yields 100 hits in the two-component library. **(c)** Chemical structures of all selected by V-SYNTHES and synthesized compounds for ROCK1 kinase.

Extended Data Fig. 14 Experimental characterization of candidate ROCK1 inhibitors predicted by V-SYNTHES.

Full dose-response curves for the ROCK1 hits in **(a)** functional potency and **(b)** binding affinity at human ROCK1. The data points are presented as mean \pm SEM from $n = 3$ independent experiments, each run carried out in triplicate. **(c)** Values of binding affinities and functional potencies for all candidate compounds predicted by V-SYNTHES. Bold font highlight hits with $IC_{50} < 10 \mu M$. Estimated values for curves that did not allow accurate fitting are marked with *.

Extended Data Fig. 15 Examples of typical Enamine REAL reactions.

(a) two-component reaction **(b)** three-component reaction.

Extended Data Table 1 Potencies and affinities of V-SYNTHES hits in functional and binding assays at CB₁ and CB₂ receptors

Supplementary information

Supplementary Information

Supplementary Figs. 1–4 and Supplementary Tables 1–4.

Reporting Summary

Supplementary File 2

Detailed synthesis protocol for all compounds in the paper.

Supplementary File 3

NMR and LC–MS spectra for all compounds in the paper.

Supplementary File 4

HRMS spectra for all compounds in the paper.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Sadybekov, A.A., Sadybekov, A.V., Liu, Y. *et al.* Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **601**, 452–459 (2022). <https://doi.org/10.1038/s41586-021-04220-9>

- Received: 17 February 2021
- Accepted: 08 November 2021
- Published: 15 December 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04220-9>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [A virtual drug-screening approach to conquer huge chemical libraries](#)

- Charlotte Deane
 - Maranga Mokaya

Nature (2022)

- [Screening ultra-large virtual libraries](#)

- Sarah Crunkhorn

Nature Reviews Drug Discovery (2022)

[A virtual drug-screening approach to conquer huge chemical libraries](#)

- Charlotte Deane
- Maranga Mokaya

News & Views 15 Dec 2021

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

- Article
- [Published: 22 December 2021](#)

Structure of Hsp90–Hsp70–Hop–GR reveals the Hsp90 client-loading mechanism

- [Ray Yu-Ruei Wang¹](#),
- [Chari M. Noddings](#) ORCID: [orcid.org/0000-0003-0890-9035¹](#),
- [Elaine Kirschke¹](#),
- [Alexander G. Myasnikov](#) ORCID: [orcid.org/0000-0003-2607-7121¹](#) nAff3,
- [Jill L. Johnson](#) ORCID: [orcid.org/0000-0003-3881-722X²](#) &
- [David A. Agard](#) ORCID: [orcid.org/0000-0003-3512-695X¹](#)

[Nature](#) volume 601, pages 460–464 (2022)

- 4672 Accesses
- 1 Citations
- 51 Altmetric
- [Metrics details](#)

Subjects

- [Chaperones](#)
- [Cryoelectron microscopy](#)
- [Nuclear receptors](#)
- [Protein folding](#)
- [Steroid hormones](#)

Abstract

Maintaining a healthy proteome is fundamental for the survival of all organisms¹. Integral to this are Hsp90 and Hsp70, molecular chaperones that together facilitate the folding, remodelling and maturation of the many ‘client proteins’ of Hsp90². The glucocorticoid receptor (GR) is a model client protein that is strictly dependent on Hsp90 and Hsp70 for activity^{3,4,5,6,7}. Chaperoning GR involves a cycle of inactivation by Hsp70; formation of an inactive GR–Hsp90–Hsp70–Hop ‘loading’ complex; conversion to an active GR–Hsp90–p23 ‘maturation’ complex; and subsequent GR release⁸. However, to our knowledge, a molecular understanding of this intricate chaperone cycle is lacking for any client protein. Here we report the cryo-electron microscopy structure of the GR-loading complex, in which Hsp70 loads GR onto Hsp90, uncovering the molecular basis of direct coordination by Hsp90 and Hsp70. The structure reveals two Hsp70 proteins, one of which delivers GR and the other scaffolds the Hop cochaperone. Hop interacts with all components of the complex, including GR, and poised Hsp90 for subsequent ATP hydrolysis. GR is partially unfolded and recognized through an extended binding pocket composed of Hsp90, Hsp70 and Hop, revealing the mechanism of GR loading and inactivation. Together with the GR-maturation complex structure⁹, we present a complete molecular mechanism of chaperone-dependent client remodelling, and establish general principles of client recognition, inhibition, transfer and activation.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

Subscribe

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

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

Fig. 1: GR-loading complex and the molecular basis of Hsp90–Hsp70 interactions.

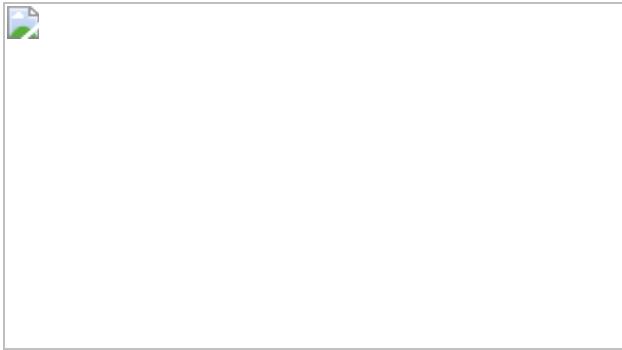


Fig. 2: Hop interacts closely with all components in the loading complex.



Fig. 3: GR is unfolded and threaded through the Hsp90 lumen, binding Hop_{DP2} and Hsp70C_{SBD-β}.

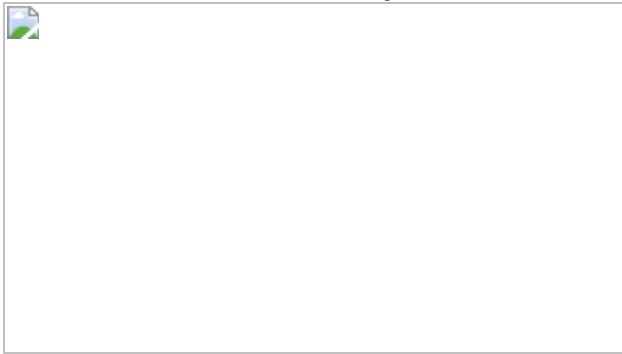
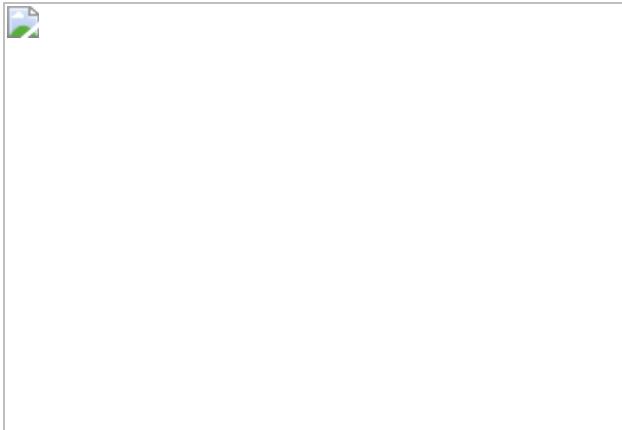


Fig. 4: Model of how Hsp70 loads GR onto Hsp90.



Data availability

The electron microscopy maps and atomic model have been deposited into the Electron Microscopy Data Bank (EMDB) and the PDB. The accession codes for the GR-loading complex are [EMD-23050](#) and [7KW7](#). Focused maps used for model refinements were also deposited with accession codes denoted in Supplementary Table 2 ([EMD-23051](#), [EMD-23053](#), [EMD-23054](#), [EMD-23055](#) and [EMD-23056](#)).

References

1. 1.

Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Hartl, F. U. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).

2. 2.

Genest, O., Wickner, S. & Doyle, S. M. Hsp90 and Hsp70 chaperones: collaborators in protein remodeling. *J. Biol. Chem.* **294**, 2109–2120 (2019).

3. 3.

Lorenz, O. R. et al. Modulation of the Hsp90 chaperone cycle by a stringent client protein. *Mol. Cell* **53**, 941–953 (2014).

4. 4.

Picard, D. et al. Reduced levels of hsp90 compromise steroid receptor action in vivo. *Nature* **348**, 166–168 (1990).

5. 5.

Nathan, D. F., Vos, M. H. & Lindquist, S. In vivo functions of the *Saccharomyces cerevisiae* Hsp90 chaperone. *Proc Natl Acad. Sci. USA* **94**, 12949–12956 (1997).

6. 6.

Smith, D. F. & Toft, D. O. Minireview: the intersection of steroid receptors with molecular chaperones: observations and questions. *Mol. Endocrinol.* **22**, 2229–2240 (2008).

7. 7.

Pratt, W. B., Morishima, Y., Murphy, M. & Harrell, M. in *Molecular Chaperones in Health and Disease* (eds Starke K. & Gaestel M.) 111–138 (2006).

8. 8.

Kirschke, E., Goswami, D., Southworth, D., Griffin, P. R. & Agard, D. A. Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. *Cell* **157**, 1685–1697 (2014).

9. 9.

Noddings, C., Wang, R. Y.-R., Johnson, J. L. & Agard, D. A. Structure of Hsp90–p23–GR reveals the Hsp90 client-remodelling mechanism. *Nature* <https://doi.org/10.1038/s41586-021-04236-1> (2021).

10. 10.

Rosenzweig, R., Nillegoda, N. B., Mayer, M. P. & Bukau, B. The Hsp70 chaperone network. *Nat. Rev. Mol. Cell Biol.* **20**, 665–680 (2019).

11. 11.

Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat. Rev. Mol. Cell Biol.* **11**, 515–528 (2010).

12. 12.

Whitesell, L. & Lindquist, S. L. HSP90 and the chaperoning of cancer. *Nat. Rev. Cancer* **5**, 761–772 (2005).

13. 13.

Lackie, R. E. et al. The Hsp70/Hsp90 chaperone machinery in neurodegenerative diseases. *Front. Neurosci.* **11**, 254 (2017).

14. 14.

Mayer, M. P. & Giersch, L. M. Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones. *J. Biol. Chem.* **294**, 2085–2097 (2019).

15. 15.

Krukenberg, K. A., Street, T. O., Lavery, L. A. & Agard, D. A. Conformational dynamics of the molecular chaperone Hsp90. *Q. Rev. Biophys.* **44**, 229–255 (2011).

16. 16.

Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nat. Rev. Mol. Cell Biol.* **18**, 345–360 (2017).

17. 17.

Boysen, M., Kityk, R. & Mayer, M. P. Hsp70- and Hsp90-mediated regulation of the conformation of p53 DNA binding domain and p53 cancer variants. *Mol. Cell* **74**, 831–843 (2019).

18. 18.

Dahiya, V. et al. Coordinated conformational processing of the tumor suppressor protein p53 by the Hsp70 and Hsp90 chaperone machineries. *Mol. Cell* **74**, 816–830 (2019).

19. 19.

Moran Luengo, T., Kityk, R., Mayer, M. P. & Rudiger, S. G. D. Hsp90 breaks the deadlock of the Hsp70 chaperone system. *Mol. Cell* **70**, 545–552 (2018).

20. 20.

Morgner, N. et al. Hsp70 forms antiparallel dimers stabilized by post-translational modifications to position clients for transfer to Hsp90. *Cell Rep.* **11**, 759–769 (2015).

21. 21.

Nathan, D. F. & Lindquist, S. Mutational analysis of Hsp90 function: interactions with a steroid receptor and a protein kinase. *Mol. Cell. Biol.* **15**, 3917–3925 (1995).

22. 22.

Bohen, S. P. & Yamamoto, K. R. Isolation of Hsp90 mutants by screening for decreased steroid receptor function. *Proc. Natl Acad. Sci. USA* **90**, 11424–11428 (1993).

23. 23.

Kravats, A. N. et al. Functional and physical interaction between yeast Hsp90 and Hsp70. *Proc. Natl Acad. Sci. USA* **115**, E2210–E2219 (2018).

24. 24.

Doyle, S. M. et al. Intermolecular interactions between Hsp90 and Hsp70. *J. Mol. Biol.* **431**, 2729–2746 (2019).

25. 25.

Genest, O., Hoskins, J. R., Kravats, A. N., Doyle, S. M. & Wickner, S. Hsp70 and Hsp90 of *E. coli* directly interact for collaboration in protein remodeling. *J. Mol. Biol.* **427**, 3877–3889 (2015).

26. 26.

Flynn, J. M. et al. Comprehensive fitness maps of Hsp90 show widespread environmental dependence. *eLife* **9**, e58310 (2020).

27. 27.

Genest, O., Hoskins, J. R., Camberg, J. L., Doyle, S. M. & Wickner, S. Heat shock protein 90 from *Escherichia coli* collaborates with the DnaK chaperone system in client protein remodeling. *Proc. Natl Acad. Sci. USA* **108**, 8206–8211 (2011).

28. 28.

Sung, N. et al. 2.4 Å resolution crystal structure of human TRAP1NM, the Hsp90 paralog in the mitochondrial matrix. *Acta Crystallogr. D* **72**, 904–911 (2016).

29. 29.

Sun, M., Kotler, J. L. M., Liu, S. & Street, T. O. The endoplasmic reticulum (ER) chaperones BiP and Grp94 selectively associate when BiP is in the ADP conformation. *J. Biol. Chem.* **294**, 6387–6396 (2019).

30. 30.

Kirschke, E., Roe-Zurz, Z., Noddings, C. & Agard, D. The interplay between Bag-1, Hsp70, and Hsp90 reveals that inhibiting Hsp70 rebinding is essential for glucocorticoid receptor activity. Preprint at <https://doi.org/10.1101/2020.05.03.075523> (2020).

31. 31.

Mandal, A. K. et al. Hsp110 chaperones control client fate determination in the hsp70-Hsp90 chaperone system. *Mol. Biol. Cell* **21**, 1439–1448 (2010).

32. 32.

Sahasrabudhe, P., Rohrberg, J., Biebl, M. M., Rutz, D. A. & Buchner, J. The plasticity of the Hsp90 co-chaperone System. *Mol. Cell* **67**, 947–961 (2017).

33. 33.

Schmid, A. B. et al. The architecture of functional modules in the Hsp90 co-chaperone Sti1/Hop. *EMBO J.* **31**, 1506–1517 (2012).

34. 34.

Southworth, D. R. & Agard, D. A. Client-loading conformation of the Hsp90 molecular chaperone revealed in the cryo-EM structure of the human Hsp90:Hop complex. *Mol. Cell* **42**, 771–781 (2011).

35. 35.

Lee, C. T., Graf, C., Mayer, F. J., Richter, S. M. & Mayer, M. P. Dynamics of the regulation of Hsp90 by the co-chaperone Sti1. *EMBO J.* **31**, 1518–1528 (2012).

36. 36.

Reidy, M., Kumar, S., Anderson, D. E. & Masison, D. C. Dual roles for yeast Sti1/Hop in regulating the Hsp90 chaperone cycle. *Genetics* **209**, 1139–1154 (2018).

37. 37.

Verba, K. A. et al. Atomic structure of Hsp90–Cdc37–Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* **352**, 1542–1547 (2016).

38. 38.

Suren, T. et al. Single-molecule force spectroscopy reveals folding steps associated with hormone binding and activation of the glucocorticoid receptor. *Proc. Natl Acad. Sci. USA* **115**, 11688–11693 (2018).

39. 39.

Bohen, S. P. Hsp90 mutants disrupt glucocorticoid receptor ligand binding and destabilize aporeceptor complexes. *J. Biol. Chem.* **270**, 29433–29438 (1995).

40. 40.

Hawle, P. et al. The middle domain of Hsp90 acts as a discriminator between different types of client proteins. *Mol. Cell. Biol.* **26**, 8385–8395 (2006).

41. 41.

Genest, O. et al. Uncovering a region of heat shock protein 90 important for client binding in *E. coli* and chaperone function in yeast. *Mol. Cell* **49**, 464–473 (2013).

42. 42.

Liu, Y. et al. Cryo-EM structures reveal a multistep mechanism of Hsp90 activation by co-chaperone Aha1. Preprint at <https://doi.org/10.1101/2020.06.30.180695> (2020).

43. 43.

Rutz, D. A. et al. A switch point in the molecular chaperone Hsp90 responding to client interaction. *Nat. Commun.* **9**, 1472 (2018).

44. 44.

Li, C. et al. FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol.* **11**, 92 (2011).

45. 45.

Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 11020–11024 (2002).

46. 46.

Obermann, W. M., Sondermann, H., Russo, A. A., Pavletich, N. P. & Hartl, F. U. In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. *J. Cell Biol.* **143**, 901–910 (1998).

47. 47.

Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).

48. 48.

Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

49. 49.

Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

50. 50.
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
51. 51.
- Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
52. 52.
- Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**, e11182 (2015).
53. 53.
- Leman, J. K. et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **17**, 665–680 (2020).
54. 54.
- Li, J. et al. Structure insights into mechanisms of ATP hydrolysis and the activation of human heat-shock protein 90. *Acta Biochim Biophys Sin.* **44**, 300–306 (2012).
55. 55.
- Arakawa, A., Handa, N., Shirouzu, M. & Yokoyama, S. Biochemical and structural studies on the high affinity of Hsp70 for ADP. *Protein Sci.* **20**, 1367–1379 (2011).
56. 56.
- Zhang, P., Leu, J. I., Murphy, M. E., George, D. L. & Marmorstein, R. Crystal structure of the stress-inducible human heat shock protein 70 substrate-binding domain in complex with peptide substrate. *PLoS ONE* **9**, e103518 (2014).

57. 57.

Wang, R. Y. et al. De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Methods* **12**, 335–338 (2015).

58. 58.

Schneider, M. et al. BiPPred: combined sequence- and structure-based prediction of peptide binding to the Hsp70 chaperone BiP. *Proteins* **84**, 1390–1407 (2016).

59. 59.

Gutierrez, M. B. B., Bonorino, C. B. C. & Rigo, M. M. ChaperISM: improved chaperone binding prediction using position-independent scoring matrices. *Bioinformatics* **36**, 735–741 (2020).

60. 60.

Zahn, M. et al. Structural studies on the forward and reverse binding modes of peptides to the chaperone DnaK. *J. Mol. Biol.* **425**, 2463–2479 (2013).

61. 61.

Soding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).

62. 62.

Song, Y. et al. High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).

63. 63.

Darby, J. F. et al. Solution structure of the Hop TPR2A domain and investigation of target druggability by NMR, biochemical and in silico approaches. *Sci. Rep.* **10**, 16000 (2020).

64. 64.

DiMaio, F., Zhang, J., Chiu, W. & Baker, D. Cryo-EM model validation using independent map reconstructions. *Protein Sci.* **22**, 865–868 (2013).

65. 65.

Wang, R. Y. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife* **5**, e17219 (2016).

66. 66.

Bledsoe, R. K. et al. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **110**, 93–105 (2002).

67. 67.

Liu, X. D., Morano, K. A. & Thiele, D. J. The yeast Hsp110 family member, Sse1, is an Hsp90 cochaperone. *J. Biol. Chem.* **274**, 26654–26660 (1999).

68. 68.

Sondermann, H. et al. Structure of a Bag/Hsc70 complex: convergent functional evolution of Hsp70 nucleotide exchange factors. *Science* **291**, 1553–1557 (2001).

69. 69.

Polier, S., Dragovic, Z., Hartl, F. U. & Bracher, A. Structural basis for the cooperation of Hsp70 and Hsp110 chaperones in protein folding. *Cell* **133**, 1068–1079 (2008).

70. 70.

Johnson, J. L., Halas, A. & Flom, G. Nucleotide-dependent interaction of *Saccharomyces cerevisiae* Hsp90 with the cochaperone proteins

Sti1, Cpr6, and Sba1. *Mol. Cell. Biol.* **27**, 768–776 (2007).

71. 71.

Flom, G., Weekes, J., Williams, J. J. & Johnson, J. L. Effect of mutation of the tetratricopeptide repeat and asparatate-proline 2 domains of Sti1 on Hsp90 signaling and interaction in *Saccharomyces cerevisiae*. *Genetics* **172**, 41–51 (2006).

72. 72.

Street, T. O. et al. Elucidating the mechanism of substrate recognition by the bacterial Hsp90 molecular chaperone. *J. Mol. Biol.* **426**, 2393–2404 (2014).

73. 73.

Kityk, R., Kopp, J. & Mayer, M. P. Molecular mechanism of J-domain-triggered ATP hydrolysis by Hsp70 chaperones. *Mol. Cell* **69**, 227–237 e224 (2018).

74. 74.

Dey, B., Lightbody, J. J. & Boschelli, F. CDC37 is required for p60v-src activity in yeast. *Mol. Biol. Cell* **7**, 1405–1417 (1996).

75. 75.

Kityk, R., Kopp, J., Sinning, I. & Mayer, M. P. Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. *Mol. Cell* **48**, 863–874 (2012).

Acknowledgements

We thank members of the Agard laboratory for discussions; T. W. Owens for advising on the photoreactive cross-linking experiment; D. Bulkley, G. Gilbert, E. Tse and Z. Yu from the W.M. Keck Foundation Advanced Microscopy Laboratory at the University of California San Francisco

(UCSF) for maintaining the electron microscopy facility and helping with data collection; and M. Harrington and J. Baker-LePain for computational support with the UCSF Wynton cluster. R.Y.-R.W. thanks D. Elnatan for various support in biochemistry at the initial stage of the project. R.Y.-R.W. was a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. C.M.N. is a National Cancer Institute Ruth L. Kirschstein Predoctoral Individual NRSA Fellow. The work was supported by funding from Howard Hughes Medical Institute (D.A.A.) and NIH grants R35GM118099 (D.A.A.), S10OD020054 (D.A.A.), S10OD021741 (D.A.A.), P20GM104420 (J.L.J.) and R01GM127675 (J.L.J.).

Author information

Author notes

1. Alexander G. Myasnikov

Present address: Dubochet Center for Imaging (DCI) at EPFL, EPFL SB IPHYS DCI, Lausanne, Switzerland

Affiliations

1. Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, USA

Ray Yu-Ruei Wang, Chari M. Noddings, Elaine Kirschke, Alexander G. Myasnikov & David A. Agard

2. Department of Biological Sciences, University of Idaho, Moscow, ID, USA

Jill L. Johnson

Contributions

R.Y.-R.W. performed the research and drafted the manuscript. D.A.A. supervised the research. C.M.N. performed pBpa-GR construct design and

purification, and helped analyse data. J.L.J. carried out all *in vivo* yeast experiments. E.K. trained R.Y.-R.W. for the biochemistry of the GR reconstitution system. A.G.M. trained R.Y.-R.W. for cryo-EM operation and data acquisition. R.Y.-R.W., C.M.N. and D.A.A. wrote the manuscript with input from all authors.

Corresponding author

Correspondence to [David A. Agard](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Oscar Llorca, Matthias Mayer 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 Purification of the GR-loading complex and the cryo-EM single-particle image processing pipeline.](#)

a, Domain organization of the chaperone proteins in the GR-loading complex. **b**, Top, elution profile of gel filtration using SEC-MALS to confirm the homogeneity of the GR-loading complex. The apparent

molecular weight of the eluent estimated by SEC-MALS is ~370 kDa, although the two-Hsp70 client-loading complex is ~440 kDa. The discrepancy may be a result of multiple species co-eluted. Bottom, SDS-PAGE stained with Coomassie blue of the eluted fractions marked in (top). **c**, SDS-PAGE stained with Coomassie blue of the fractions treated with 0.02% (w/v) glutaraldehyde cross-linking for 20 min at room temperature, followed by quenching with 20 mM Tris buffer at pH 7.5. Data in (**b-c**) are representative data of at least two independent experiments. **d**, Initial model generation for the GR-loading complex. The 60 Å low-pass filtered initial model used to reconstruct the 3D model was adopted from the Hsp90 semi-open conformation structure from the Hsp90:Hop cryo-EM structure³⁴. **e**, Schematic workflow of the global cryo-EM map reconstruction. Yellow boxes indicate the selected class to move forward. Blue box indicates one-Hsp70 loading complex. Purple box indicates the final high-resolution global reconstruction. **f**, Flow chart of focused classification/refinement using the signal subtraction approach from RELION. Final reconstructions for individual masked classifications/refinements were selected based on the resolution intercepted with the FSC 0.143 from 3D auto-refine.

K=number of classes; *T*=regularization factor, Tau.

Extended Data Fig. 2 Resolution estimates for cryo-EM reconstructions, atomic B-factor refinement and model-map FSC.

a, Local resolution estimates for the GR-loading complex global reconstruction were calculated using RELION with front view (left) and back view (right). **b**, Euler angle distribution in the final reconstruction. Orthogonal views of the reconstruction are shown with front view (top) and side view (bottom) **c**, Gold-standard FSC for the global cryo-EM reconstruction. **d**, Atomic model with B-factors refined with colour key shown on the left. **e**, Histogram of the B-factor values of all non-hydrogen atoms in the atomic model, coloured by the same colour key in **d**. **f**, Model-map FSC. **g**, Focused classification and refinement of the GR-loading complex. Masks were created at various regions of the GR-loading complex (left) and its corresponding gold-standard FSC (right) after 3D auto-refine. The nominal resolution for each reconstruction is labelled and indicated in the FSC plots.

Extended Data Fig. 3 The Hsp90 in both two-Hsp70 and one-Hsp70 GR-loading complexes adopts a semi-closed conformation.

a, Close-up view of the novel dimerization interface of the symmetric Hsp90 dimer with residues at the interface in transparent surface representation. The interface is composed of two molecular switches of Hsp90, the first helix and the lid motif. **b, c**, The Hsp90 in the loading complex (**b**) is one step away from the fully closed ATP state (**c**). Front (left) and side (right) views of the Hsp90 in the two-Hsp90 GR-loading complex (**b**) and the Hsp90 in the GR-maturation complex⁹ (**c**). Arrows indicate displacements from the Hsp90 in the loading state, in which a large twisting motion is apparent from the side view. **d**, Comparison of cryo-EM reconstructions of the one-Hsp70 and two-Hsp70 GR-loading complexes. Schematic model of the two-Hsp70 loading complex (left). Front views of cryo-EM maps of the two-Hsp70 (middle; grey colour) and one-Hsp70 (right; salmon colour) GR-loading complexes. Right, the one-Hsp70 reconstruction has lost density for Hsp70C_{NBD} (dashed circle) and Hsp70C_{SBD- α} (red arrows); however, density for GR_{pre-Helix 1} and GR_{Helix 1} (black arrows) is in the same location as it is in the two-Hsp70 GR-loading complex. **e**, Rigid-body fitting of the two Hsp90 protomers individually in the one- or two-Hsp70 loading complexes cryo-EM reconstructions shows both of the Hsp90s share a similar semi-closed conformation. The Hsp90 in the one-Hsp70 loading complex (middle panel) has a slightly wider opening angle (right panel) than the Hsp90 in the two-Hsp70 loading complex (left panel). **f**, The lumen of the semi-closed Hsp90 presented in the loading complex can fit a helix⁷². A helix (magenta) can be accommodated in the semi-closed Hsp90. Front view (left) and top view (right).

Extended Data Fig. 4 Interfaces I and II are both crucial for Hsp90–Hsp70 interactions and client activation.

a, b, The Hsp70 cleft (dashed circles), formed by Hsp70_{NBD-IA} and Hsp70_{NBD-IIA} (**a**), which Hsp90_{MD} interacts with, is used by the Hsp70 interdomain linker in the Hsp70^{ATP} state and Hsp40's J-domain (**b**).

Hsp90A:Hsp70C in the loading complex are shown as cartoon (**a**, left) and surface (**a**, right) representation. The *E. coli* Hsp70 (eHsp70):J-protein complex in the ATP state (PDB ID: 5NRO²³) are shown as cartoon (**b**, left) and surface (**b**, right) representation. The two subdomains of eHsp70 are coloured in green for eHsp70^{ATP}_{NBD} and in pink for eHsp70^{ATP}_{SBD}. The *E. coli* Hsp40 J-domain (eHsp40_{J-domain}) is coloured in purple. **c, d**, Mapping Hsp90/Hsp70 residues previously characterized by the Wickner group on the loading complex. Five yeast Hsp90 (Hsp82^{P281,G313,K394,K398,K399})²³ and four yeast Hsp70 (Ssa1^{R169,N172,E210,T219})²⁴ mutations previously characterized by the Wickner group^{23,24} are all located at Interface I. These ‘Wickner residues’ are shown with stick and transparent sphere representation, whereas residues that interact with the Wickner residues are shown with only stick representation. Polar interactions are highlighted with green dashed lines. Residue numbers of the Wickner residues in yeast Hsp90/Hsp70 are shown as the labels within parentheses. Based on the proximity of the Wickner residue positions to Interface I residues, our structure can readily explain why Hsp82^{P281C} (P301, Hsp90) and Ssa1^{T219C(T222, Hsp70)} (**d**) showed no effect, whereas the other four Hsp90 mutants (Hsp82^{G313S(G333, Hsp90)}, Hsp82^{K394C(K414, Hsp90)}, Hsp82^{K398E(K418, Hsp90)}, Hsp82^{K399C(K419, Hsp90)}) and the other three Hsp70 mutants (Ssa1^{N172D(N174, Hsp70)}, Ssa1^{R169H(R171, Hsp70)}, Ssa1^{E210R(D213, Hsp70)}) (**c**) disrupted Hsp90:Hsp70 interaction significantly. **e–h**, In vivo validation of Interface I and II of the Hsp90:Hsp70 interactions in the GR-loading complex. Mapping the positions of the three mutations (arrows and yellow surface representation in **e**) used for in vivo validation on the atomic structure of Hsp90:Hsp70 (dark blue:dark orange; surface representation in **e**) in the GR-loading complex. The Interface I residues used, Hsp90^{G333} (Hsc82^{G309}) and Hsp90^{K418} (Hsc82^{K394}), are indicated by arrows (**e**, left panel). The Interface II residue, Hsp90^{R60} (Hsc82^{R46}), is indicated by an arrow (**e**, right panel). Residue numbers of the residues in yeast Hsc82 are shown as the labels within parentheses in **e**. In **f**, His-Hsc82 complexes were isolated from yeast and analysed by SDS-PAGE and visualized by Coomassie staining and immunoblot analysis (see also Supplementary Fig. 10 for the uncropped gels/blots). Yeast proteins: St1=Hop; Ssa1/2=Hsp70; Hsc82=Hsp90β. In **g**, plasmids expressing wild-type or mutant Hsc82 were

expressed as the sole Hsp90 in JJ816 (*hsc82hsp82*) cells. Growth was examined by spotting 10-fold serial dilutions of yeast cultures on rich media, followed by incubation for two days at 30 °C or 37 °C. In **h**, strains expressing wild-type (WT) or mutant *HSC82* were transformed with a multicopy plasmid expressing *GAL1-v-src* (pBv-src) or the control plasmid (pB656)⁷⁴. Yeast cultures were grown overnight at 30 °C in raffinose-uracil drop-out medium until mid-log phase. Galactose (20%) was added to a final concentration of 2%. After six hours, cultures were serially diluted 10-fold onto uracil drop-out plates containing galactose. Plates were grown for 2-3 days at 30 °C.

Extended Data Fig. 5 Both the Hsp90^{ATP} and the Hsp70^{ATP} conformations are incompatible with the GR-loading complex.

a, Overlay of the crystal structure of Apo Hsp90 fragment (purple; PDB ID: 3T0H⁵⁴) to the Hsp90A_{NTD} (dark blue). Green circle highlights the open ATP pocket lid. **b**, Closure of the ATP pocket lid in the ATP state of Hsp90_{NTD} (the Hsp90α structure from the GR-maturation complex⁹ is in yellow, ribbon representation) clashes (magenta circle) with the Hsp70_{NBD} (orange, surface and ribbon representation) in the loading complex. The NTD fragment of Hsp90^{ATP} is aligned with the Hsp90_{NTD} in the loading complex. **c**, Superimposition of the ATP state of Hsp90_{NTD-MD} fragment (yellow) to the Hsp90A (dark blue) at the MD. Magenta circles indicate steric clashes of the ATP state of the Hsp90_{NTD} to the Hsp70_{NBD} (orange; surface/ribbon representation). **d**, Superposition of the Hsp70^{ATP} conformation (green; PDB ID: 4B9Q⁷⁵) to the Hsp70C_{NBD} (dark orange). Arrows indicate the two subdomains of Hsp70^{ATP}_{SBD}, which cause serious steric clashes with the Hsp90 in the loading complex shown in **e**. **e**, The superimposed Hsp70^{ATP} shown in **d** is fixed from **d** and the Hsp90 (dark blue; surface/ribbon representation) of the GR-loading complex is present. Magenta circles highlight steric clashes caused by the two subdomains of Hsp70^{ATP}_{SBD} (green) to the Hsp90A_{NTD-MD}.

Extended Data Fig. 6 The Hsp70–Hop interface in the GR-loading complex is crucial for cellular functions and client maturation.

a–f, The atomic interactions of Hsp70S_{NBD}:Hop_{TPR2A}:Hsp90B_{MEEVD} in the GR-loading complex. The cryo-EM map from focused classification and refinement is shown in (**a** and **d**, left). The atomic model with the corresponding view from (**a** and **d**, left) is shown in the right panels of (**a**, **d**). Close-up views of the Hsp70S_{NTD}:Hop_{TPR2A} interface with the atomic model fit into the density (**b**). Two key residues (Hop^{Y296} and Hop^{A328}) are buried in the surface with their sidechain density indicated by the arrows (**b**, top and bottom). Density of the interface residues highlighted in Fig. 2b is shown in (**b**, bottom). Sequence alignments of Hop with key residues at the Hsp70_{NBD}:Hop_{TPR2A} interface highlighted by the red triangles in (**c**), in which the colour scheme is BLOSUM62. Close-up view of the atomic model of Hop_{TPR2A}:Hsp90B_{MEEVD} fit with the cryo-EM map is shown in (**e**). Close-up view of the atomic interactions of the MEEVD fragment from Hsp90B (light blue) and Hop_{TPR2A} (pink) from **e** is shown in **f**, in which polar interactions are depicted with dashed lines. **g–j**, In vivo validation of the Hsp70S_{NBD}:Hop_{TPR2A} interface in the loading complex. The two buried residues in Hsp70S_{NBD-IIA}:Hop_{TPR2A} interface, which were chosen for the mutational studies, are shown in **g**. Components of the GR-loading complex (**g**, top left) are coloured as in other figures and as labelled. Hsp70S_{NBD} is shown in surface-charge representation (blue: positive; red: negative) calculated using PyMOL. Note that the corresponding residue numbering of the two positions in yeast Hop (Sti1) are shown in parentheses of the labels of the bottom panels in **g**. In **h**, Sti1^{Y332A-T364H} accumulates at levels similar to WT Sti1 (see also Supplementary Fig. 10 for the uncropped gels/blots); data in **h** are representative of two independent experiments. Extracts from WT cells (JJ762), *stil* cells (JJ623), or *stil* cells transformed with a plasmid that expresses WT Sti1 or Sti1^{Y332A-T364H} were analysed by SDS–PAGE and immunoblotted with a polyclonal antisera specific for Sti1. Loading control is antibody against mitochondrial protein Tim44. In **i**, *stil-Y332A-T364H* is inviable in *hsc82hsp82* cells expressing *hsc82-G309S*. *hsc82hsp82* (JJ117) or *stilhsc82hsp82* (JJ1443) strains harbouring YEp24-

HSP82 were transformed with plasmids expressing WT *HSC82* or *hsc82-G309S*. Strains that lacked *STII* were also transformed with an empty plasmid or a plasmid expressing WT *STII* or *stil-Y332A-T364H*. Transformants were grown in the presence of 5-FOA for 3 days to counter-select for the YEp24-*HSP82* plasmid. *STII* is essential under these conditions and the growth of cells expressing *stil-Y332A-T364H* was indistinguishable from those expressing the empty plasmid. In **j**, WT cells, *stil* cells or *stil* cells transformed with a plasmid that expresses WT Stil or Stil-Y332A-T364H were transformed with an empty plasmid or a plasmid that expresses GAL-v-src. v-src induction in the presence of galactose sharply reduces the growth of WT cells, but not cells lacking *STII*. The growth of cells expressing *stil-Y332A-T364H* was very similar to those expressing the empty plasmid, indicating that *stil-Y332A-T364H* is unable to support v-src function. The growth of cells in the presence of glucose was indistinguishable. 10-fold serial dilutions of cultures were grown for 3 days in the presence of galactose or glucose.

Extended Data Fig. 7 The cryo-EM density and atomic model of GRHelix 1 motif interacting with Hsp90 and HopDP2.

a, The focused map of the Hsp90AB_{CTD}:Hsp70C_{SBD-β}:Hop^{DP2}:GR_{Helix 1} (top) and the atomic model shown in ribbon representation (bottom). **b**, The top view of the reconstruction and model shown in **a**. **c**, The density (mesh) for the GR_{Helix 1} motif (residues 528-551) gripped by Hsp90 and Hop_{DP2}. **d**, The atomic interactions of the GR_{Helix 1} motif with Hsp90 and Hop_{DP2} corresponding to **b**, cyan box in bottom). Residues in contact with the GR motif are shown in stick representation. The types of molecular interaction Hsp90 and Hop_{DP2} provide are indicated at the top; H and P denote hydrophobic and polar interactions, respectively. **e**, The 7 Å low-pass filtered cryo-EM map shows that the lumen density (yellow shade) of GR connects to the globular part of GR on the other side of Hsp90. **f**, Docking of the GRLBD to the 10 Å low-pass filtered map shows that the low-resolution GR density can fit the rest of the GRLBD. **g**, The low-pass filtered map shows that W320 and F349 (arrows) of Hsp90A in the loading complex are in contact with GRLBD.

Extended Data Fig. 8 Validations of Hop_{DP2} binding to GR and of the in vivo importance of the client-binding pocket in Hop_{DP2}.

a–d, Using photoreactive, site-directed cross-linking to validate the Hop_{DP2}:GR interaction (**a**, the pink box) and (**b**, left). In **a**, Hop_{DP2} is loosely packed and uses surface-exposed hydrophobic residues, shown in sticks, to interact with Hsp90A_{amphi-α} (shown in transparent surface and with hydrophobic residues in sticks) and Hsp90B^{W320,F349}. Suggested by modelling of the photoreactive cross-linker *p*-benzoyl-l-phenylalanine (pBpa) on various positions of Hop_{DP2} to search for the position which is at the closest proximity to the GR_{Helix1} density (**b**, right), the pBpa was placed at Hop^{Q512} (**b**, left). The blue arrow on the right panel in **b** points at the selected position, Q512 (right panel, **b**). A time course of UV-exposed GR-loading complex analysed by SDS–PAGE and visualized by Coomassie staining (**c**). Whole fractions of GR-loading complex eluted from the size-exclusion column were exposed to UV using a gel imager (see also Methods). In **c**, arrows at 0 and 60 min indicate a reduced intensity of the GR band over the time course. Western blot of the SDS–PAGE gel after a 60 min UV exposure, using anti-MBP antibody to detect the MBP-tagged GR is shown in (**d**). Data in (**c, d**) are from one experiment. **e–h**, Hop_{DP2}'s client-binding/transfer function is crucial for cellular functions and client maturation. Hop^{L508} (L553 in Sti1) is located on the hydrophobic palm (Fig. 2d) of Hop^{DP2}, interacting closely with the LXXLL motif of GR_{Helix 1} through hydrophobic interactions (**e**, left, middle and right). Mutations of Hop^{L508} completely abrogated GR function in vivo (Sti1^{L553A} in Schmid et al. 2012³³), lead to growth defects (**g**), and failed to promote v-src maturation (**h**). The mutant Sti1^{L553D} accumulates at levels similar to WT Sti1 (**f**); data in **f** are from two independent experiments (see also Supplementary Fig. 10 for the uncropped gels/blots). Extracts from WT cells (JJ762), sti1 cells (JJ623) or sti1 cells transformed with a plasmid that expresses WT Sti1 or Sti1^{L553D} were analysed by SDS–PAGE and immunoblotted with a polyclonal antisera specific for Sti1. Loading control is antibody against mitochondrial protein Tim44. In **g**, *sti1-L553D* is

inviable in *hsc82hsp82* cells expressing *hsc82-G309S*. *hsc82hsp82* (JJ117) or *stilhsc82hsp82* (JJ1443) strains harbouring YEp24-*HSP82* were transformed with plasmids expressing WT *HSC82* or *hsc82-G309S*. Strains that lacked *STIL* were also transformed with an empty plasmid or a plasmid expressing WT *STIL* or *stil-L553D*. Transformants were grown in the presence of 5-FOA for 3 days to counter-select for the YEp24-*HSP82* plasmid. *STIL* is essential under these conditions and the growth of cells expressing *stil-L553D* was indistinguishable from those expressing the empty plasmid. In **h**, WT cells, *stil* cells or *stil* cells transformed with a plasmid that expresses WT Stil or Stil-L553D were transformed with an empty plasmid or a plasmid that expresses GAL-v-src. v-src induction in the presence of galactose sharply reduces the growth of WT cells, but not cells lacking *STIL*. The growth of cells expressing *stil-L553D* was very similar to those expressing the empty plasmid, indicating that *stil-L553D* is unable to support v-src function. The growth of cells in the presence of glucose was indistinguishable. 10-fold serial dilutions of cultures were grown for 3 days in the presence of galactose or glucose.

Extended Data Fig. 9 Hsp70 inhibits GR by binding the pre-Helix 1 region of GR.

a, After engaging with Hsp70/Hsp40, the GR_{pre-Helix 1} region exhibits protection from deuterium incorporation in a HD-exchange mass spectrometry (HDX-MS) experiment⁸. In the GR_{pre-Helix 1} region, there are Hsp70-binding sites predicted by two state-of-the-art algorithms (BiP Pred⁵⁸ and ChaperISM⁵⁹). **b**, Left, GRLBD crystal structure (PDB ID: 1M2Z) coloured by the change of deuterium uptake (HDX-MS data was retrieved from a previous study)⁸; green: protection from deuterium incorporation; yellow: positive deuterium uptake. Right, the detachment of the entire GR_{Helix 1} motif explains the positive deuterium uptake around the ligand-binding pocket. **c**, The protection from deuterium incorporation of the GR_{pre-Helix 1} region can be explained by the binding of Hsp70. Together **b** and **c** provide a molecular mechanism describing how Hsp70 can inhibit GR ligand binding. **d**, GR's pre-Helix 1 remains bound to Hsp70C_{SBD-β} (red circles) in the loading complex. Left, the 6Å low-pass filtered cryo-EM map of the loading complex. Right, atomic model with ribbon presentation.

Extended Data Fig. 10 The GR_{pre-Helix 1} strand is engaged with the client-loading Hsp70 (Hsp70C) in the GR-loading complex.

a, The photoreactive cross-linker (pBpa) was placed at two positions in the GR_{pre-Helix 1} strand, residues before (GR^{P517}, sphere presentation) and after (GR^{Q527}, sphere representation) the predicted Hsp70-binding site (GR^{519–526}, stick representation). **b, c**, Whole fractions of GR-loading complex eluted from the size-exclusion column were exposed to UV using a gel imager (see also Methods). As expected, at both positions, cross-links between GR and Hsp70 were formed in the GR-loading complex, indicated by high molecular weight species. Left, SDS-PAGE stained with Coomassie blue. Middle, anti-Hsp70 Western blot. Right, anti-Hop Western blot. Data from **(b, c)** are from one experiment. Hsp70 cross-linking efficiency was higher for the GR^{pBpa517} position (**b**, middle) than for the GR^{pBpa527} position (**c**, middle). It is likely because 1) the C-terminal end of the Hsp70-bound substrate tends to be flexible in the reverse binding mode, as indicated by high atomic B-factors and missing density from previously determined Hsp70 crystal structures with a reverse peptide bound (PDB ID: 4EZZ, 4EZQ, 4EZT, and 4EZY)⁶⁰, and 2) GR⁵²⁷ is closer to Hop⁵⁴³ than Hsp70. In addition, the two positions were able to cross-link with Hop (**b, c**, right), indicating it is the client-loading Hsp70 that the GR_{pre-Helix 1} strand cross-linked with, rather than the scaffolding Hsp70. Note that although the GR⁵¹⁷ position is not adjacent to Hop in the GR-loading complex model, we reasoned that the cross-link may be formed in the one-Hsp70 loading complex (**b**, right). These results demonstrate that it is the GR_{pre-Helix 1} strand bound to Hsp70C, supporting our structural model. **d, e**, Raw western blots from middle and right panels in **b, c**. Red pixels in the western blots shown indicate overexposure.

Supplementary information

Supplementary Information

This file contains Supplementary Figs 1–10, additional references and Supplementary Tables 1–3.

Reporting Summary

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Wang, R.YR., Noddings, C.M., Kirschke, E. *et al.* Structure of Hsp90–Hsp70–Hop–GR reveals the Hsp90 client-loading mechanism. *Nature* **601**, 460–464 (2022). <https://doi.org/10.1038/s41586-021-04252-1>

- Received: 02 December 2020
- Accepted: 16 November 2021
- Published: 22 December 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04252-1>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [**Structure of Hsp90-p23-GR reveals the Hsp90 client-remodelling mechanism**](#)
 - Chari M. Noddings
 - Ray Yu-Ruei Wang
 - David A. Agard

Nature (2021)

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04252-1>

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

- Article
- [Published: 22 December 2021](#)

Structure of Hsp90–p23–GR reveals the Hsp90 client-remodelling mechanism

- [Chari M. Noddings](#) ORCID: orcid.org/0000-0003-0890-9035¹ na1,
- [Ray Yu-Ruei Wang](#)¹ na1,
- [Jill L. Johnson](#) ORCID: orcid.org/0000-0003-3881-722X² &
- [David A. Agard](#) ORCID: orcid.org/0000-0003-3512-695X¹

Nature volume 601, pages 465–469 (2022)

- 3406 Accesses
- 1 Citations
- 45 Altmetric
- [Metrics details](#)

Subjects

- [Chaperones](#)
- [Cryoelectron microscopy](#)
- [Nuclear receptors](#)
- [Protein folding](#)
- [Steroid hormones](#)

Abstract

Hsp90 is a conserved and essential molecular chaperone responsible for the folding and activation of hundreds of ‘client’ proteins^{1,2,3}. The glucocorticoid receptor (GR) is a model client that constantly depends on Hsp90 for activity^{4,5,6,7,8,9}. GR ligand binding was previously shown to be inhibited by Hsp70 and restored by Hsp90, aided by the co-chaperone p23¹⁰. However, a molecular understanding of the chaperone-mediated remodelling that occurs between the inactive Hsp70–Hsp90 ‘client-loading complex’ and an activated Hsp90–p23 ‘client-maturation complex’ is lacking for any client, including GR. Here we present a cryo-electron microscopy (cryo-EM) structure of the human GR-maturation complex (GR–Hsp90–p23), revealing that the GR ligand-binding domain is restored to a folded, ligand-bound conformation, while being simultaneously threaded through the Hsp90 lumen. In addition, p23 directly stabilizes native GR using a C-terminal helix, resulting in enhanced ligand binding. This structure of a client bound to Hsp90 in a native conformation contrasts sharply with the unfolded kinase–Hsp90 structure¹¹. Thus, aided by direct co-chaperone–client interactions, Hsp90 can directly dictate client-specific folding outcomes. Together with the GR-loading complex structure¹², we present the molecular mechanism of chaperone-mediated GR remodelling, establishing the first, to our knowledge, complete chaperone cycle for any Hsp90 client.

This is a preview of subscription content

Access options

Subscribe to nature+

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

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

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

Fig. 1: Architecture of the GR-maturation complex.

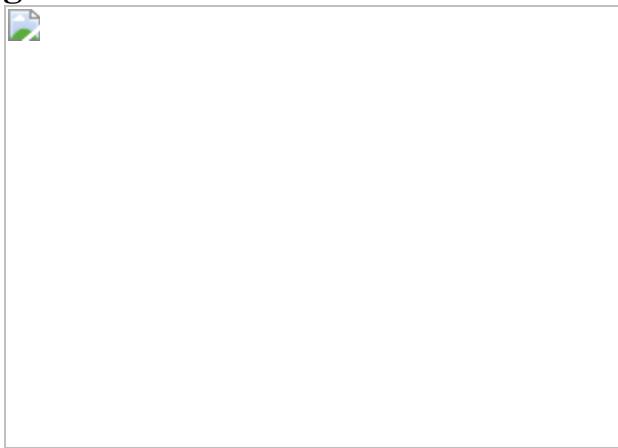


Fig. 2: p23 tail helix interactions and effect on GR ligand binding.

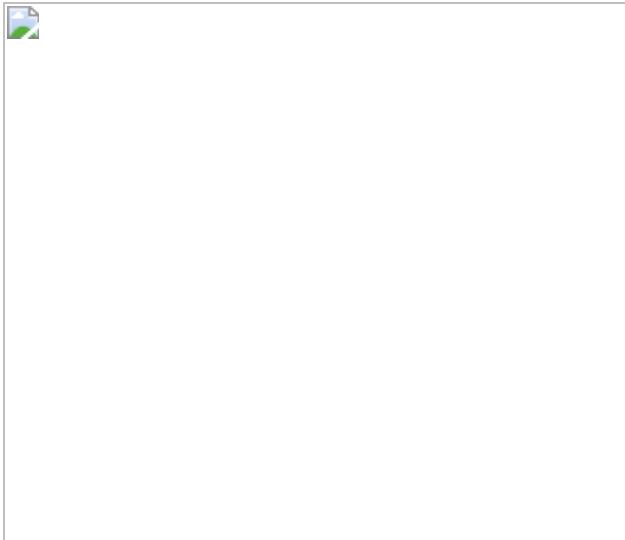
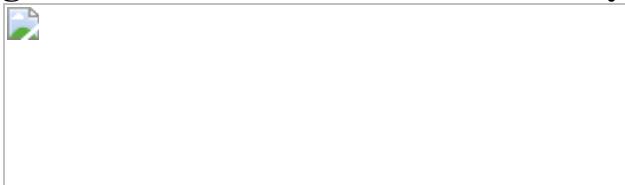


Fig. 3: Mechanism of GR activation by Hsp90.



Data availability

The cryo-EM maps generated in this study have been deposited in the Electron Microscopy Data Bank (EMDB) under the accession codes [EMD-23004](#) (GR–Hsp90–p23), [EMD-23006](#) (Hsp90–p23) and [EMD-23005](#) (MBP–Hsp90–p23). The atomic coordinates have been deposited in the PDB under the accession code [7KJL](#) (GR–Hsp90–p23). Publicly available PDB entries used in this study are: 5FWK, 1M2Z, 4P6X, 1EJF, 2CG9, 1OMP and 1ANF. The human p23 structure prediction is available from AlphaFold v2.0 with the accession code [P83868](#). [Source data](#) are provided with this paper.

References

1. 1.

Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat. Rev. Mol. Cell Biol.*

11, 515–528 (2010).

2. 2.

Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nat. Rev. Mol. Cell Biol.* **18**, 345–360 (2017).

3. 3.

Taipale, M. et al. Quantitative analysis of HSP90–client interactions reveals principles of substrate recognition. *Cell* **150**, 987–1001 (2012).

4. 4.

Picard, D. et al. Reduced levels of hsp90 compromise steroid receptor action in vivo. *Nature* **348**, 166–168 (1990).

5. 5.

Pratt, W. B. & Toft, D. O. Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocrine Rev.* **18**, 306–360 (1997).

6. 6.

Morishima, Y., Murphy, P. J., Li, D. P., Sanchez, E. R. & Pratt, W. B. Stepwise assembly of a glucocorticoid receptor hsp90 heterocomplex resolves two sequential ATP-dependent events involving first hsp70 and then hsp90 in opening of the steroid binding pocket. *J. Biol. Chem.* **275**, 18054–18060 (2000).

7. 7.

Smith, D. F. & Toft, D. O. Minireview: the intersection of steroid receptors with molecular chaperones: observations and questions. *Mol. Endocrinol.* **22**, 2229–2240 (2008).

8. 8.

Lorenz, O. R. et al. Modulation of the Hsp90 chaperone cycle by a stringent client protein. **53**, 941–953 (2014).

9. 9.

Nathan, D. F. & Lindquist, S. Mutational analysis of Hsp90 function: interactions with a steroid receptor and a protein kinase. *Mol. Cell. Biol.* **15**, 3917–3925 (1995).

10. 10.

Kirschke, E., Goswami, D., Southworth, D., Griffin, P. & Agard, D. Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. *Cell* **157**, 1685–1697 (2014).

11. 11.

Verba, K. A. et al. Atomic structure of Hsp90–Cdc37–Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* **352**, 1542–1547 (2016).

12. 12.

Wang, R. Y.-R. et al. Structure of Hsp90–Hsp70–Hop–GR reveals the Hsp90 client-loading mechanism. *Nature* <https://doi.org/10.1038/s41586-021-04252-1> (2021).

13. 13.

Zhao, R. et al. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* **120**, 715–727 (2005).

14. 14.

Rosenzweig, R., Nillegoda, N. B., Mayer, M. P. & Bukau, B. The Hsp70 chaperone network. *Nat. Rev. Mol. Cell Biol.* **20**, 665–680 (2019).

15. 15.

Krukenberg, K. A., Street, T. O., Lavery, L. A. & Agard, D. A. Conformational dynamics of the molecular chaperone Hsp90. *Q. Rev. Biophys.* **44**, 229–255 (2011).

16. 16.

Ali, M. M. U. et al. Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 (2006).

17. 17.

Sahasrabudhe, P., Rohrberg, J., Biebl, M. M., Rutz, D. A. & Buchner, J. The plasticity of the Hsp90 co-chaperone system. *Mol. Cell* **67**, 947–961.e945 (2017).

18. 18.

Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

19. 19.

Wang, R. Y. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).

20. 20.

Meyer, P. et al. Structural and functional analysis of the middle segment of Hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol. Cell* **11**, 647–658 (2003).

21. 21.

Rutz, D. A. et al. A switch point in the molecular chaperone Hsp90 responding to client interaction. *Nat. Commun.* **9**, 1472 (2018).

22. 22.

Hawle, P. et al. The middle domain of Hsp90 acts as a discriminator between different types of client proteins. *Mol. Cell. Biol.* **26**, 8385–8395 (2006).

23. 23.

Bledsoe, R. K. et al. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Nature* **419**, 93–105 (2002).

24. 24.

Weikl, T., Abelmann, K. & Buchner, J. An unstructured C-terminal region of the Hsp90 co-chaperone p23 is important for its chaperone function. *J. Mol. Biol.* **293**, 685–691 (1999).

25. 25.

Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

26. 26.

Seraphim, T. V. et al. The C-terminal region of the human p23 chaperone modulates its structure and function. *Arch. Biochem. Biophys.* **565**, 57–67 (2015).

27. 27.

Biebl, M. M. et al. Structural elements in the flexible tail of the co-chaperone p23 coordinate client binding and progression of the Hsp90 chaperone cycle. *Nat. Commun.* **12**, 828 (2021).

28. 28.

de Castro, E. et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365 (2006).

29. 29.

McKenna, N. J. & O'Malley, B. W. Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108**, 465–474 (2002).

30. 30.

Weaver, A. J., Sullivan, W. P., Felts, S. J., Owen, B. A. L. & Toft, D. O. Crystal structure and activity of human p23, a heat shock protein 90 co-chaperone. *J. Biol. Chem.* **275**, 23045–23052 (2000).

31. 31.

Freeman, B. C., Toft, D. O. & Morimoto, R. I. Molecular chaperone machines: chaperone activities of the cyclophilin Cyp-40 and the steroid aporeceptor-associated protein p23. *Science* **274**, 1718–1720 (1996).

32. 32.

Freeman, B. C., Felts, S. J., Toft, D. O. & Yamamoto, K. R. The p23 molecular chaperones act at a late step in intracellular receptor action to differentially affect ligand efficacies. *Genes Dev.* **14**, 422–434 (2000).

33. 33.

Bohen, S. P. Genetic and biochemical analysis of p23 and ansamycin antibiotics in the function of Hsp90-dependent signaling proteins. *Mol. Cell. Biol.* **18**, 3330–3339 (1998).

34. 34.

Freeman, B. C. & Yamamoto, K. R. Disassembly of transcriptional regulatory complexes by molecular chaperones. *Science* **296**, 2232–2235 (2002).

35. 35.

Liu, Y., Elnatan, D., Sun, M., Myasnikov, A. G. & Agard, D. A. Cryo-EM reveals the dynamic interplay between mitochondrial Hsp90 and SdhB folding intermediates. Preprint at <https://doi.org/10.1101/2020.10.06.327627> (2020).

36. 36.

Suren, T. et al. Single-molecule force spectroscopy reveals folding steps associated with hormone binding and activation of the glucocorticoid receptor. *Proc. Natl Acad. Sci. USA* **115**, 11688–11693 (2018).

37. 37.

Czar, M. J., Galigniana, M. D., Silverstein, A. M. & Pratt, W. B. Geldanamycin, a heat shock protein 90-binding benzoquinone ansamycin, inhibits steroid-dependent translocation of the glucocorticoid receptor from the cytoplasm to the nucleus. *Biochemistry* **36**, 7776–7785 (1997).

38. 38.

Galigniana, M. D., Radanyi, C., Renoir, J. M., Housley, P. R. & Pratt, W. B. Evidence that the peptidylprolyl isomerase domain of the hsp90-binding immunophilin FKBP52 is involved in both dynein interaction and glucocorticoid receptor movement to the nucleus. *J. Biol. Chem.* **276**, 14884–14889 (2001).

39. 39.

Netzer, W. J. & Hartl, F. U. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**, 343–349 (1997).

40. 40.

Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612

(2004).

41. 41.

Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

42. 42.

Johnson, J. L. & Toft, D. O. Binding of p23 and hsp90 during assembly with the progesterone receptor. *Mol. Endocrinol.* **9**, 670–678 (1995).

43. 43.

Csermely, P. et al. Atp induces a conformational change of the 90-Kda heat-shock protein (Hsp90). *J. Biol. Chem.* **268**, 1901–1907 (1993).

44. 44.

Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).

45. 45.

Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

46. 46.

Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

47. 47.

Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

48. 48.

Song, Y. et al. High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).

49. 49.

Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **31**, 10657–10663 (1992).

50. 50.

Quiocho, F. A., Spurlino, J. C. & Rodseth, L. E. Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **5**, 997–1015 (1997).

51. 51.

UniProt, C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

52. 52.

Madeira, F. et al. The EMBL–EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

53. 53.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

54. 54.

Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

55. 55.

Landau, M. et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).

56. 56.

Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

57. 57.

He, Y. et al. Structures and mechanism for the design of highly potent glucocorticoids. *Cell Res.* **24**, 713–726 (2014).

58. 58.

Pei, J. & Grishin, N. V. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700–712 (2001).

59. 59.

Mirabello, C. & Pollastri, G. Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics* **29**, 2056–2058 (2013).

60. 60.

Kallberg, M. et al. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012).

61. 61.

Johnson, J. L., Halas, A. & Flom, G. Nucleotide-dependent interaction of *Saccharomyces cerevisiae* Hsp90 with the cochaperone proteins Sti1, Cpr6, and Sba1. *Mol. Cell. Biol.* **27**, 768–776 (2007).

62. 62.

Johnson, J. L. & Craig, E. A. A role for the Hsp40 Ydj1 in repression of basal steroid receptor activity in yeast. *Mol. Cell. Biol.* **20**, 3027–3036 (2000).

Acknowledgements

We thank members of the Agard laboratory and E. Kirschke for helpful discussions; D. Bulkley, G. Gilbert, Z. Yu and E. Tse from the W. M. Keck Foundation Advanced Microscopy Laboratory at the University of California, San Francisco (UCSF) for EM facility maintenance and help with data collection; and M. Harrington and J. Baker-LePain for computational support with the UCSF Wynton cluster. C.M.N. is a National Cancer Institute Ruth L. Kirschstein Predoctoral Individual NRSA Fellow. R.Y.-R.W. was a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. The work was supported by funding from Howard Hughes Medical Institute (D.A.A.) and NIH grants R35GM118099 (D.A.A.), S10OD020054 (D.A.A.), S10OD021741 (D.A.A.), P20GM104420 (J.L.J.) and R01GM127675 (J.L.J.).

Author information

Author notes

1. These authors contributed equally: Chari M. Noddings, Ray Yu-Ruei Wang

Affiliations

1. Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA

Chari M. Noddings, Ray Yu-Ruei Wang & David A. Agard

2. Department of Biological Sciences, University of Idaho, Moscow, ID, USA

Jill L. Johnson

Contributions

C.M.N. and R.Y.-R.W. designed and executed biochemical experiments, cryo-EM sample preparation, data collection, data processing and model building. J.L.J. executed yeast *in vivo* assays and interpreted the results. C.M.N., R.Y.-R.W. and D.A.A. conceived the project, interpreted the results and wrote the manuscript.

Corresponding author

Correspondence to [David A. Agard](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Matthias Mayer 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 Sample Preparation.

a, Coomassie-stained raw, uncropped SDS-PAGE (4-12% acrylamide gel) with elution from the MBP-GR pulldown from the *in vitro* reconstituted GR chaperone cycle. Assay conditions are as follows- Lane 1: 5 μ M MBP-GR only; Lane 2-5: 5 μ M MBP-GR, 2 μ M Hsp40, 5 μ M Hsp70, 5 μ M Hop, 15 μ M Hsp90, 15 μ M Bag-1, 30 μ M p23, 5 mM ATP. Lane 2: no Molybdate addition, Lane 3: addition of 20mM Molybdate, Lane 4: addition of 20 mM Molybdate after 1 h pre-incubation (see [Methods](#)) **b**, Shodex KW-804 size exclusion chromatography profile of the GR-maturation complex purified by MBP-GR pulldown from the reconstituted GR chaperone cycle. mAU=milli-absorbance units. **c**, Coomassie-stained raw, uncropped SDS-PAGE (4-12% acrylamide gel) of the fractions from size exclusion chromatography. Colors indicate which gel lanes correspond to specific regions of the size exclusion chromatography profile. Sample fractions from the region highlighted in purple were collected and used for cryo-EM data collection. This experiment was repeated 4 independent times with similar results. **d**, Representative electron micrograph for the cryo-EM dataset ($-2.32 \mu\text{m}$ defocus). A total of 5608 micrographs were obtained. Scale bar is 20nm. **e**, Domain organization of the proteins in the GR-maturation complex.

Extended Data Fig. 2 Cryo-EM Data Analysis.

a, Cryo-EM data processing procedure with 3D reconstructions colored by local resolution. **b**, Gold-standard Fourier shell correlation (FSC) curves of the 3D reconstructions. The black dashed lines intercept the y-axis at an FSC value of 0.143. For the GR:Hsp90:p23 reconstruction, the map-model FSC is plotted in blue, with the blue dashed line intercepting the y-axis at an FSC value of 0.5. **c**, GR maturation complex map density with atomic model showing ATP-magnesium density in both Hsp90 protomers (Hsp90A/B). Bottom images show increased contour level on the map density to indicate that the ATP γ -phosphate position has much stronger

density relative to the α and β -phosphates, likely corresponding to molybdate, which may act as a γ -phosphate analog (see [Methods](#)).

Extended Data Fig. 3 Hsp90:GR Interfaces.

Atomic model of the maturation complex with Hsp90A (dark blue), Hsp90B (light blue), GR (yellow). **a**, View of the GR_{pre-helix 1} strand threaded through the Hsp90 lumen and GR helices 1 and 3 packing against the entrance to the Hsp90 lumen. Side chains on GR in contact with Hsp90 are shown. Hsp90A/B are in surface representation. Hydrophobic residues on Hsp90 are colored in pink. **b**, Interface 1 of the Hsp90:GR interaction depicting the GR_{pre-Helix 1} region (GR^{523–531}) threading through the Hsp90 lumen. Side chains in contact between GR and Hsp90 are shown, along with hydrogen bonds (dashed pink lines). **c**, Interface 2 of the Hsp90:GR interaction depicting GR_{Helix 1} (GR^{532–539}) packing against Hsp90B. Side chains in contact between GR and Hsp90 are shown, along with hydrogen bonds (dashed pink lines). **d**, Interface 3 of the Hsp90:GR interaction depicting residues on the Hsp90A_{MD} loops (Hsp90A^{N318,W320,F349,R346}) and Hsp90B_{amphi- α} (Hsp90B^{T624,Y627,M628}) packing against GR. Side chains in contact between GR and Hsp90 are shown, along with hydrogen bonds (dashed pink lines).

Extended Data Fig. 4 GR is in a Native, Ligand-Bound State in the Maturation Complex.

a, Atomic model of GR from the maturation complex (yellow) compared with GR from the crystal structure (PDB ID 1M2Z) (light pink) with co-activator peptide NCoA2 (purple) and ligand (pink). GR_{Helix 12} is indicated. **b**, GR-maturation complex atomic model in surface representation with the co-activator peptide NCoA2 (purple) docked based on the GR:NCoA2 crystal structure (PDB ID M2Z). The NCoA2 peptide binding interface is available and the bound NCoA2 peptide does not clash with Hsp90. Hsp90A (dark blue), Hsp90B (light blue), GR (yellow), p23 (green). **c**, GR maturation complex map density (sharpened with B factor –40) with either the dexamethasone-bound crystal structure docked (left panel, PDB ID

1M2Z) or the cortisol-bound crystal structure docked (right panel, PDB ID 4P6X) into the GR map density. In the top images, the ligand density is shown with the agonist dexamethasone (left) or the agonist cortisol (right) from the docked crystal structures. Arrow indicates the extra carbon atom in dexamethasone compared to cortisol. In the bottom images, density for GR^{Y735} is shown with either the dexamethasone-bound crystal structure docked (left) or the cortisol-bound crystal structure docked (right). **d**, GR-maturation complex atomic model in surface representation depicting the GR LBD dimerization interface. Hsp90A (dark blue), Hsp90B (light blue), GR (yellow), p23 (green). Left, the GR LBD dimerization interface is highlighted (light pink). Right, while the dimerization interface is solvent accessible in the GR-maturation complex, the binding of a second GR LBD (light pink) clashes with the Hsp90B CTD. The dimerization interface is based on the GR LBD dimer crystal structure (PDB ID 1M2Z).

Extended Data Fig. 5 Hsp90:p23 Interfaces.

Atomic model of the maturation complex with Hsp90A (dark blue), Hsp90B (light blue), GR (yellow), p23 (green). **b**, Interface 1 of the Hsp90:p23 interaction depicting Hsp90B interacting with one side of the p23 core. Side chains in contact between p23 and Hsp90B are shown, along with hydrogen bonds (dashed pink lines). **c**, Interface 2 of the Hsp90:p23 interaction depicting Hsp90B interacting with the base of the p23 core. Side chains in contact between p23 and Hsp90B are shown, along with hydrogen bonds (dashed pink lines). **d**, Interface 3 of the Hsp90:p23 interaction depicting Hsp90A interacting with the side of the p23 core. Side chains in contact between p23 and Hsp90A are shown, along with hydrogen bonds (dashed pink lines). **e**, Atomic model of a symmetric Hsp90 dimer (orange) compared with Hsp90 from the maturation complex atomic model, indicating a slight asymmetry in the Hsp90 dimer interface in the maturation complex. Hsp90A (dark blue), Hsp90B (light blue), p23 (green).

Extended Data Fig. 6 The p23tail helix:GR Interface.

a, Focused map of GR:p23_{tail helix} showing density for the p23 tail with the atomic model built in. GR (yellow), p23 (green). **b**, Interface between the

p23_{tail} helix (green) and GR (colored by hydrophobicity, surface representation) showing that the p23_{tail} helix binds to a hydrophobic patch on GR. p23 side chains interacting with GR are shown. **c**, Sequence identity across human SHRs (GR, mineralocorticoid receptor, androgen receptor, progesterone receptor, estrogen receptor α and β) plotted onto the GR structure. The p23_{tail} helix (light green) was overlaid to indicate the p23:GR interface. **d**, Secondary structure predictions for human p23 from four different servers. Porter 4.0 (orange), RaptorX (blue), Psipred (purple), AlphaFold v2.0 (pink). The p23_{tail} helix from the maturation complex atomic model is shown with the top green line. **e**, Atomic model of GR (yellow) and p23 (green) from the maturation complex highlighting the interaction between the p23_{tail} helix and the GR C-terminus, which connects to GR_{Helix 12}. **f**, Sequence alignment of eukaryotic p23 showing conservation of the p23_{tail} helix sequence. The p23_{tail} helix from the maturation complex atomic model is shown with the top green line. The bottom aligned sequence is the p23_{tail} helix -like motif identified in NCoA3 using the ScanProsite server. Red boxes on the *S. cerevisiae* p23 sequence indicate predicted helices from the PsiPred server. The alignment is colored according to the ClustalW convention.

Extended Data Fig. 7 Effect of p23 Tail Mutants on GR Activity and Cell Survival.

a, Depiction of the two p23 tail mutants used in the GR activity assays. **b**, Individual data points corresponding to Fig. 2d. Equilibrium binding of 20 nM fluorescent dexamethasone to 250 nM GR with chaperones and p23 tail mutants measured by fluorescence polarization (mean±SD). n=3 biologically independent samples per condition (n=6 biologically independent samples for the GR only condition). Significance was evaluated using a one-way ANOVA ($F_{(3,8)} = 636.2$; $p < 0.0001$) with *post-hoc* Dunnett's multiple comparisons test (n.s. $P \geq 0.05$; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$). P-values: $p(p23 \text{ vs. } p23\Delta\text{tail}) = 0.1512$, $p(p23 \text{ vs. } p23\Delta\text{helix}\Delta\text{tail}) = 0.0002$, $p(p23 \text{ vs. no p23}) = <0.0001$. **c**, Equilibrium binding of 20 nM fluorescent dexamethasone to 250 nM GR with addition of 15 μM p23 or p23 tail mutants measured by fluorescence

polarization (mean \pm SD). n=7 biologically independent samples per condition (n=6 biologically independent samples for the GR + p23 Δ tail condition). Fluorescence polarization values are baseline subtracted in accordance with the measured fluorescent dexamethasone baseline polarization value. There were no statistically significant differences between group means as determined by a one-way ANOVA ($F_{(3,23)} = 1.708$; $p=0.1933$). **d**, Yeast survival assay with human p23 or p23 tail mutants. Top panels: *hsc82hsp82Δ* yeast expressing Hsc82 I588A-M589A exhibit a growth defect at 37 °C in the presence of *SBA1* and enhanced defects in cells lacking *SBA1* (*sba1*). Bottom panels: Growth is restored by addition of human p23, although p23 Δ helix Δ tail exhibits reproducibly reduced growth relative to p23 or p23 Δ tail. **e**, GR activation assay in wild-type yeast strain JJ762 expressing p23, p23 mutants, or Sba1 in addition to wild-type amounts of Sba1 from the native promoter. The fold increase in GR activities compared to the empty vector (e.v.) control are shown (mean \pm SD). n=18 biologically independent samples per condition (10 independent samples for the +Sba1 condition). Significance was evaluated using a one-way ANOVA ($F_{(4,77)} = 7.077$; $p < 0.0001$) with *post-hoc* Šídák's multiple comparisons test (n.s. $P \geq 0.05$; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$). P-values: $p(\text{e.v. vs. p23}) = 0.0001$, $p(\text{p23 vs p23}\Delta\text{tail}) = 0.0164$, $p(\text{p23 vs. p23}\Delta\text{helix}\Delta\text{tail}) = 0.0021$, $p(\text{p23 vs. Sba1}) = 0.0002$, $p(\text{p23}\Delta\text{tail vs. p23}\Delta\text{helix}\Delta\text{tail}) = 0.9721$. **f**, Expression of human p23 or p23 tail mutants in wild-type yeast strain JJ762 assayed by immunoblot with polyclonal antisera raised against Sba1 or human p23.

[Source data](#)

[**Extended Data Fig. 8 Hsp90:p23 Complex.**](#)

a, Cryo-EM density map of the Hsp90:p23 complex. Hsp90A (dark blue), Hsp90B (light blue), p23 (green). This color scheme is maintained in all figures that show the structure. **b**, Atomic model of Hsp90 and p23 from the GR-maturation complex docked into the Hsp90:p23 map density. **c**, Top view of the Hsp90:p23 complex density map with clipping plane to show unidentified density (gray) through the Hsp90 lumen. **d**, Cartoon representation of the Hsp90:p23 complex illustrating that MBP and GR LBD are not present in the map density (represented by a gray box).

Extended Data Fig. 9 MBP:Hsp90:p23 Complex.

a, Cryo-EM density map of the MBP:Hsp90:p23 complex. Far right image shows the density map lowpass-filtered to 8Å. Hsp90A (dark blue), Hsp90B (light blue), p23 (green), MBP (orange). This color scheme is maintained throughout **b**, Apo MBP crystal structure (PDB ID 1OMP) and atomic model of Hsp90 and p23 from the GR-maturation complex docked into the MBP:Hsp90:p23 map density. Note the missing density for the two MBP C-terminal helices. Far right image shows the density map lowpass-filtered to 8Å. **c**, Maltose-bound MBP crystal structure (PDB ID 1ANF) docked into the MBP:Hsp90:p23 map density. MBP (orange), maltose (pink). **d**, Top view of the MBP:Hsp90:p23 complex density map with clipping plane to show unidentified density (gray) through the Hsp90 lumen. **e**, Cartoon representation of the MBP:Hsp90:p23 complex illustrating the GR LBD is not present in the map density (represented by a gray box).

Extended Data Fig. 10 Comparison of the GR-Maturation Complex with the Hsp90:Kinase Complex.

a, Structure of Hsp90 bound to an unfolded kinase client (PDB ID 5FWK) with a strand of the kinase client threaded through the Hsp90 lumen. The two hydrophobic residues on the kinase (Cdk4^{V89,V92}) that occupy the Hsp90 hydrophobic pockets are displayed. In the GR-maturation complex, two hydrophobic residues on GR (GR^{L525,L528}) occupy the Hsp90 hydrophobic pockets, demonstrating a conserved client binding mode. Hsp90A (dark blue, surface representation), Hsp90B (light blue, surface representation), Cdk4 kinase (purple). Hydrophobic residues on Hsp90 are colored in pink. **b**, Structure of Hsp90 bound to an unfolded kinase client (PDB ID 5FWK) depicting Hsp90A^{F341} (Hsp90 isoform β) and Hsp90B_{amphi-α} packing against the kinase. In the GR-maturation complex, the corresponding residue Hsp90A^{F349} (Hsp90 isoform α) and the Hsp90B_{amphi-α} also pack against GR, demonstrating a conserved Hsp90:client binding interface. Hsp90A (dark blue), Hsp90B (light blue), Cdk4 kinase (purple, surface representation). **c**, Top, atomic models of the GR-maturation complex and Hsp90:kinase complex showing that both

clients thread through the closed Hsp90 lumen. Hsp90A (dark blue, surface representation), Hsp90B (light blue, surface representation), GR (yellow), Cdk4 kinase (purple). Bottom, schematics demonstrating that both clients thread through the mostly hydrophobic Hsp90 lumen, but have different folding outcomes (H=hydrophobic interface).

Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics

Supplementary information

Reporting Summary

Peer Review File

Source data

Source Data Fig. 2

Source Data Extended Data Fig. 7

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Noddings, C.M., Wang, R.YR., Johnson, J.L. *et al.* Structure of Hsp90–p23–GR reveals the Hsp90 client-remodelling mechanism. *Nature* **601**, 465–469 (2022). <https://doi.org/10.1038/s41586-021-04236-1>

- Received: 02 December 2020

- Accepted: 13 November 2021
- Published: 22 December 2021
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04236-1>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [**Structure of Hsp90âHsp70âHopâGR reveals the Hsp90 client-loading mechanism**](#)
 - Ray Yu-Ruei Wang
 - Chari M. Noddings
 - David A. Agard

Nature (2021)

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04236-1>

Amendments & Corrections

- [**Author Correction: In situ formation of ZnOx species for efficient propane dehydrogenation**](#) [06 January 2022]
Author Correction •
- [**Author Correction: A neuroanatomical basis for electroacupuncture to drive the vagal–adrenal axis**](#) [06 January 2022]
Author Correction •
- [**Author Correction: A human-specific modifier of cortical connectivity and circuit function**](#) [06 January 2022]
Author Correction •
- [**Author Correction: Cryo-EM structure of the Ebola virus nucleoprotein–RNA complex at 3.6 Å resolution**](#) [06 January 2022]
Author Correction •
- [**Author Correction: Human neocortical expansion involves glutamatergic neuron diversification**](#) [06 January 2022]
Author Correction •
- [**Publisher Correction: Observation of Stark many-body localization without disorder**](#) [06 January 2022]
Publisher Correction •

- Author Correction
- Open Access
- [Published: 06 January 2022](#)

Author Correction: In situ formation of ZnO_x species for efficient propane dehydrogenation

- [Dan Zhao](#) ORCID: orcid.org/0000-0002-8301-7654^{1,2},
- [Xinxin Tian](#) ORCID: orcid.org/0000-0002-3195-3176^{2,3},
- [Dmitry E. Doronkin](#)⁴,
- [Shanlei Han](#) ORCID: orcid.org/0000-0002-6110-0203^{1,2},
- [Vita A. Kondratenko](#)²,
- [Jan-Dierk Grunwaldt](#) ORCID: orcid.org/0000-0003-3606-0956⁴,
- [Anna Perechodjuk](#)²,
- [Thanh Huyen Vuong](#) ORCID: orcid.org/0000-0001-5535-7064²,
- [Jabor Rabeh](#) ORCID: orcid.org/0000-0003-2162-0981²,
- [Reinhard Eckelt](#)²,
- [Uwe Rodemerck](#)²,
- [David Linke](#) ORCID: orcid.org/0000-0002-5898-1820²,
- [Guixuan Jiang](#) ORCID: orcid.org/0000-0003-1464-3368¹,
- [Haijun Jiao](#) ORCID: orcid.org/0000-0002-2947-5937² &
- [Evgenii V. Kondratenko](#) ORCID: orcid.org/0000-0003-0431-6937²

[Nature](#) volume 601, page E8 (2022)

- 807 Accesses
- 3 Altmetric

- [Metrics details](#)

Subjects

- [Heterogeneous catalysis](#)
- [Materials for energy and catalysis](#)
- [Process chemistry](#)

The [Original Article](#) was published on 10 November 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03923-3>
Published online 10 November 2021

In the version of this Article initially published, an error appeared in Extended Data Fig. 1a,b, where the content of the panels (plots and units, not axis labels) were inadvertently switched. The corrected version, with black data points in panel a, and blue, red and black data points in panel b, now appears correctly online.

The original Article has been corrected online.

Author information

Affiliations

1. State Key Laboratory of Heavy Oil Processing, China University of Petroleum, Beijing, P. R. China

Dan Zhao, Shanlei Han & Guiyuan Jiang

2. Leibniz-Institut für Katalyse e.V., Rostock, Germany

Dan Zhao, Xinxin Tian, Shanlei Han, Vita A. Kondratenko, Anna Perechodjuk, Thanh Huyen Vuong, Jabor Rabeah, Reinhard

Eckelt, Uwe Rodemerck, David Linke, Haijun Jiao & Evgenii V. Kondratenko

3. Key Laboratory of Materials for Energy Conversion and Storage of Shanxi Province, Institute of Molecular Science, Shanxi University, Taiyuan, P. R. China

Xinxin Tian

4. Institute of Catalysis Research and Technology and Institute for Chemical Technology and Polymer Chemistry, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Dmitry E. Doronkin & Jan-Dierk Grunwaldt

Corresponding authors

Correspondence to [Guixuan Jiang](#), [Haijun Jiao](#) or [Evgenii V. Kondratenko](#).

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

Zhao, D., Tian, X., Doronkin, D.E. *et al.* Author Correction: In situ formation of ZnO_x species for efficient propane dehydrogenation. *Nature* **601**, E8 (2022). <https://doi.org/10.1038/s41586-021-04285-6>

- Published: 06 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04285-6>

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/s41586-021-04285-6>

- Author Correction
- [Published: 06 January 2022](#)

Author Correction: A neuroanatomical basis for electroacupuncture to drive the vagal–adrenal axis

- [Shenbin Liu](#)^{1,2,3,4} ✉nal,
- [Zhifu Wang](#)¹ ✉nal,
- [Yangshuai Su](#)^{1,5},
- [Lu Qi](#)¹,
- [Wei Yang](#)¹,
- [Mingzhou Fu](#)¹,
- [Xianghong Jing](#)⁵,
- [Yanqing Wang](#)^{2,3,4} &
- [Qiufu Ma](#) ORCID: orcid.org/0000-0001-7896-7914¹

[Nature](#) volume 601, page E9 (2022)

- 1227 Accesses
- 4 Altmetric
- [Metrics details](#)

Subjects

- [Neural circuits](#)
- [Neurophysiology](#)

The [Original Article](#) was published on 13 October 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-04001-4> published online 13 October 2021

In the version of this article initially published, there was an omission in affiliation 4. The full affiliation has been amended to read: “State Key Laboratory of Medical Neurobiology, Institutes of Brain Science, Fudan University, Shanghai, China.”

The error has been corrected in the online version of the article.

Author information

Author notes

1. These authors contributed equally: Shenbin Liu, Zhifu Wang

Affiliations

1. Dana-Farber Cancer Institute and Department of Neurobiology, Harvard Medical School, Boston, MA, USA

Shenbin Liu, Zhifu Wang, Yangshuai Su, Lu Qi, Wei Yang, Mingzhou Fu & Qiufu Ma

2. Institute of Acupuncture and Moxibustion, Fudan Institutes of Integrative Medicine, Fudan University, Shanghai, China

Shenbin Liu & Yanqing Wang

3. Department of Integrative Medicine and Neurobiology, School of Basic Medical Science, Fudan University, Shanghai, China

Shenbin Liu & Yanqing Wang

4. Institutes of Brain Science, Fudan University, Shanghai, China

Shenbin Liu & Yanqing Wang

5. Meridians Research Center, Institute of Acupuncture and Moxibustion,
China Academy of Chinese Medical Sciences, Beijing, China

Yangshuai Su & Xianghong Jing

Corresponding author

Correspondence to [Qiufu Ma](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Liu, S., Wang, Z., Su, Y. *et al.* Author Correction: A neuroanatomical basis for electroacupuncture to drive the vagal–adrenal axis. *Nature* **601**, E9 (2022). <https://doi.org/10.1038/s41586-021-04290-9>

- Published: 06 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04290-9>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04290-9>

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

- Author Correction
- [Published: 06 January 2022](#)

Author Correction: A human-specific modifier of cortical connectivity and circuit function

- [Ewoud R. E. Schmidt](#) [ORCID: orcid.org/0000-0003-4654-2377](#)^{1,2 nAff6},
- [Hanzhi T. Zhao](#) [ORCID: orcid.org/0000-0001-6616-7039](#)^{2,3 nAff6},
- [Jung M. Park](#)^{1,2},
- [Mario Dipoppa](#)^{2,4 na1},
- [Mauro M. Monsalve-Mercado](#)^{2,4 na1},
- [Jacob B. Dahan](#)^{1,2},
- [Chris C. Rodgers](#)^{1,2,5},
- [Amélie Lejeune](#)^{1,2},
- [Elizabeth M. C. Hillman](#) [ORCID: orcid.org/0000-0001-5511-1451](#)^{2,3},
- [Kenneth D. Miller](#)^{1,2,4,5},
- [Randy M. Bruno](#)^{1,2,5} &
- [Franck Polleux](#) [ORCID: orcid.org/0000-0003-4313-0481](#)^{1,2,5}

[Nature](#) volume **601**, page E10 (2022)

- 752 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Genetics of the nervous system](#)
- [Neural circuits](#)

The [Original Article](#) was published on 27 October 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-04039-4> published online 27 October 2021

In the version of this article initially published, there was an error in ref. 23. The reference, rather than citing Park, J. *IBRO Rep.* **6** S550 (2019), should instead have cited Park, J. M. et al. Deep and superficial layers of the primary somatosensory cortex are critical for whisker-based texture discrimination in mice. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.08.12.245381> (2020).

The reference has been corrected in the online version of the article.

Author information

Author notes

1. Ewoud R. E. Schmidt & Hanzhi T. Zhao

Present address: Department of Neuroscience, Medical University of South Carolina, Charleston, SC, USA

2. These authors contributed equally: Mario Dipoppa, Mauro M. Monsalve-Mercado

Affiliations

1. Department of Neuroscience, Columbia University, New York, NY, USA

Ewoud R. E. Schmidt, Jung M. Park, Jacob B. Dahan, Chris C. Rodgers, Amélie Lejeune, Kenneth D. Miller, Randy M. Bruno & Franck Polleux

2. Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA

Ewoud R. E. Schmidt, Hanzhi T. Zhao, Jung M. Park, Mario Dipoppa, Mauro M. Monsalve-Mercado, Jacob B. Dahan, Chris C. Rodgers, Amélie Lejeune, Elizabeth M. C. Hillman, Kenneth D. Miller, Randy M. Bruno & Franck Polleux

3. Department of Biomedical Engineering and Radiology, Columbia University, New York, NY, USA

Hanzhi T. Zhao & Elizabeth M. C. Hillman

4. Center for Theoretical Neuroscience, College of Physicians and Surgeons, Columbia University, New York, NY, USA

Mario Dipoppa, Mauro M. Monsalve-Mercado & Kenneth D. Miller

5. Kavli Institute for Brain Science, Columbia University, New York, NY, USA

Chris C. Rodgers, Kenneth D. Miller, Randy M. Bruno & Franck Polleux

Corresponding author

Correspondence to [Franck Polleux](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Schmidt, E.R.E., Zhao, H.T., Park, J.M. *et al.* Author Correction: A human-specific modifier of cortical connectivity and circuit function. *Nature* **601**, E10 (2022). <https://doi.org/10.1038/s41586-021-04302-8>

- Published: 06 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04302-8>

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/s41586-021-04302-8>

- Author Correction
- [Published: 06 January 2022](#)

Author Correction: Cryo-EM structure of the Ebola virus nucleoprotein–RNA complex at 3.6 Å resolution

- [Yukihiko Sugita](#)¹ nAff7,
- [Hideyuki Matsunami](#)¹,
- [Yoshihiro Kawaoka](#)^{2,3,4},
- [Takeshi Noda](#)^{5,6} &
- [Matthias Wolf](#) ORCID: orcid.org/0000-0001-8480-0918¹

[Nature](#) volume 601, page E11 (2022)

- 1203 Accesses
- 5 Altmetric
- [Metrics details](#)

Subjects

- [Cryoelectron microscopy](#)
- [Ebola virus](#)

The [Original Article](#) was published on 17 October 2018

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-018-0630-0> published online 17 October 2018

In the version of this article initially published, there was an error in the units listed in the last paragraph of the main text. The sentence should have read “On the basis of its 6 nucleotides per NP subunit and genome size of the Zaire ebolavirus (18,960 bases), the length of the nucleocapsid necessary to accommodate a single copy of genomic RNA is 950 nm”, rather than “950 μm”.

Author information

Author notes

1. Yukihiko Sugita

Present address: Laboratory of Advanced Protein Characterization, Research Center for State-of-the-Art Functional Protein Analysis, Institute for Protein Research, Osaka University, Osaka, Japan

Affiliations

1. Molecular Cryo-Electron Microscopy Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

Yukihiko Sugita, Hideyuki Matsunami & Matthias Wolf

2. Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo, Japan

Yoshihiro Kawaoka

3. Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI, USA

Yoshihiro Kawaoka

4. Department of Special Pathogens, International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Tokyo, Japan

Yoshihiro Kawaoka

5. Laboratory of Ultrastructural Virology, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, Japan

Takeshi Noda

6. PRESTO, Japan Science and Technology Agency, Saitama, Japan

Takeshi Noda

Corresponding author

Correspondence to [Matthias Wolf](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Sugita, Y., Matsunami, H., Kawaoka, Y. *et al.* Author Correction: Cryo-EM structure of the Ebola virus nucleoprotein–RNA complex at 3.6 Å resolution. *Nature* **601**, E11 (2022). <https://doi.org/10.1038/s41586-021-04304-6>

- Published: 06 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04304-6>

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/s41586-021-04304-6>

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

- Author Correction
- Open Access
- [Published: 06 January 2022](#)

Author Correction: Human neocortical expansion involves glutamatergic neuron diversification

- [Jim Berg](#) [ORCID: orcid.org/0000-0002-3300-5399¹](#), [na1](#),
- [Staci A. Sorenson¹](#), [na1](#),
- [Jonathan T. Ting^{1,2}](#), [na1](#),
- [Jeremy A. Miller](#) [ORCID: orcid.org/0000-0003-4549-588X¹](#), [na1](#),
- [Thomas Chartrand¹](#),
- [Anatoly Buchin¹](#),
- [Trygve E. Bakken](#) [ORCID: orcid.org/0000-0003-3373-7386¹](#),
- [Agata Budzillo](#) [ORCID: orcid.org/0000-0002-2723-3272¹](#),
- [Nick Dee¹](#),
- [Song-Lin Ding¹](#),
- [Nathan W. Gouwens](#) [ORCID: orcid.org/0000-0001-8429-4090¹](#),
- [Rebecca D. Hodge](#) [ORCID: orcid.org/0000-0002-5784-9668¹](#),
- [Brian Kalmbach](#) [ORCID: orcid.org/0000-0003-3136-8097^{1,2}](#),
- [Changkyu Lee¹](#),
- [Brian R. Lee¹](#),
- [Lauren Alfiler¹](#),
- [Katherine Baker¹](#),
- [Eliza Barkan¹](#),
- [Allison Beller³](#),
- [Kyla Berry](#) [ORCID: orcid.org/0000-0002-0632-6322¹](#),

- [Darren Bertagnolli](#)¹,
- [Kris Bickley](#)¹,
- [Jasmine Bomben](#)¹,
- [Thomas Braun](#)⁴,
- [Krissy Brouner](#)¹,
- [Tamara Casper](#)¹,
- [Peter Chong](#)¹,
- [Kirsten Crichton](#)¹,
- [Rachel Dalley](#)¹,
- [Rebecca de Frates](#) [ORCID: orcid.org/0000-0002-7820-093X](#)¹,
- [Tsega Desta](#)¹,
- [Samuel Dingman Lee](#)¹,
- [Florence D’Orazi](#)¹,
- [Nadezhda Dotson](#)¹,
- [Tom Egdorf](#)¹,
- [Rachel Enstrom](#)¹,
- [Colin Farrell](#)¹,
- [David Feng](#)¹,
- [Olivia Fong](#)¹,
- [Szabina Furdan](#)⁵,
- [Anna A. Galakhova](#)⁶,
- [Clare Gamlin](#)¹,
- [Amanda Gary](#)¹,
- [Alexandra Glandon](#) [ORCID: orcid.org/0000-0002-1539-3332](#)¹,
- [Jeff Goldy](#)¹,
- [Melissa Gorham](#) [ORCID: orcid.org/0000-0002-6770-9474](#)¹,
- [Natalia A. Goriounova](#) [ORCID: orcid.org/0000-0002-5917-983X](#)⁶,
- [Sergey Gratiy](#)¹,
- [Lucas Graybuck](#) [ORCID: orcid.org/0000-0002-8814-6818](#)¹,
- [Hong Gu](#)¹,
- [Kristen Hadley](#) [ORCID: orcid.org/0000-0002-4504-8724](#)¹,
- [Nathan Hansen](#)⁷,
- [Tim S. Heistek](#)⁶,
- [Alex M. Henry](#)¹,

- [Djai B. Heyer](#) ORCID: [orcid.org/0000-0002-8118-1450⁶](http://orcid.org/0000-0002-8118-1450),
- [DiJon Hill¹](#),
- [Chris Hill¹](#),
- [Madie Hupp](#) ORCID: [orcid.org/0000-0002-7444-5286¹](http://orcid.org/0000-0002-7444-5286),
- [Tim Jarsky](#) ORCID: [orcid.org/0000-0002-4399-539X¹](http://orcid.org/0000-0002-4399-539X),
- [Sara Kebede](#) ORCID: [orcid.org/0000-0002-3775-1583¹](http://orcid.org/0000-0002-3775-1583),
- [Lisa Keene³](#),
- [Lisa Kim¹](#),
- [Mean-Hwan Kim¹](#),
- [Matthew Kroll¹](#),
- [Caitlin Latimer³](#),
- [Boaz P. Levi¹](#),
- [Katherine E. Link¹](#),
- [Matthew Mallory¹](#),
- [Rusty Mann¹](#),
- [Desiree Marshall³](#),
- [Michelle Maxwell¹](#),
- [Medea McGraw¹](#),
- [Delissa McMillen¹](#),
- [Erica Melief³](#),
- [Eline J. Mertens⁶](#),
- [Leona Mezei⁵](#),
- [Norbert Mihut](#) ORCID: [orcid.org/0000-0002-6432-3656⁵](http://orcid.org/0000-0002-6432-3656),
- [Stephanie Mok¹](#),
- [Gabor Molnar](#) ORCID: [orcid.org/0000-0001-7959-139X⁵](http://orcid.org/0000-0001-7959-139X),
- [Alice Mukora¹](#),
- [Lindsay Ng¹](#),
- [Kiet Ngo¹](#),
- [Philip R. Nicovich¹](#),
- [Julie Nyhus](#) ORCID: [orcid.org/0000-0002-2006-8235¹](http://orcid.org/0000-0002-2006-8235),
- [Gaspar Olah⁵](#),
- [Aaron Oldre](#) ORCID: [orcid.org/0000-0002-4969-8759¹](http://orcid.org/0000-0002-4969-8759),
- [Victoria Omstead¹](#),
- [Attila Ozsvár⁵](#),

- [Daniel Park¹](#),
- [Hanchuan Peng¹](#),
- [Trangthanh Pham](#) [ORCID: orcid.org/0000-0002-4738-5062¹](#),
- [Christina A. Pom¹](#),
- [Lydia Potekhina¹](#),
- [Ramkumar Rajanbabu¹](#),
- [Shea Ransford¹](#),
- [David Reid¹](#),
- [Christine Rimorin¹](#),
- [Augustin Ruiz](#) [ORCID: orcid.org/0000-0002-8361-5152¹](#),
- [David Sandman¹](#),
- [Josef Sulc¹](#),
- [Susan M. Sunkin¹](#),
- [Aaron Szafer¹](#),
- [Viktor Szemenyei⁵](#),
- [Elliot R. Thomsen](#) [ORCID: orcid.org/0000-0001-9092-8678¹](#),
- [Michael Tieu¹](#),
- [Amy Torkelson¹](#),
- [Jessica Trinh¹](#),
- [Herman Tung](#) [ORCID: orcid.org/0000-0002-0812-3318¹](#),
- [Wayne Wakeman¹](#),
- [Femke Waleboer⁶](#),
- [Katelyn Ward¹](#),
- [René Wilbers](#) [ORCID: orcid.org/0000-0002-6743-2188⁶](#),
- [Grace Williams¹](#),
- [Zizhen Yao](#) [ORCID: orcid.org/0000-0002-9361-5607¹](#),
- [Jae-Geun Yoon⁷](#),
- [Costas Anastassiou](#) [ORCID: orcid.org/0000-0002-6793-0611¹](#),
- [Anton Arkhipov¹](#),
- [Pal Barzo⁸](#),
- [Amy Bernard](#) [ORCID: orcid.org/0000-0003-2540-1153¹](#),
- [Charles Cobbs⁷](#),
- [Philip C. de Witt Hamer](#) [ORCID: orcid.org/0000-0003-2988-8544⁹](#),
- [Richard G. Ellenbogen¹⁰](#),

- [Luke Esposito¹](#),
- [Manuel Ferreira¹⁰](#),
- [Ryder P. Gwinn⁷](#),
- [Michael J. Hawrylycz](#) ORCID: orcid.org/0000-0002-5741-8024¹,
- [Patrick R. Hof¹¹](#),
- [Sander Idema⁹](#),
- [Allan R. Jones¹](#),
- [C. Dirk Keene](#) ORCID: orcid.org/0000-0002-5291-1469³,
- [Andrew L. Ko¹⁰](#),
- [Gabe J. Murphy^{1,2}](#),
- [Lydia Ng¹](#),
- [Jeffrey G. Ojemann¹⁰](#),
- [Anoop P. Patel¹⁰](#),
- [John W. Phillips¹](#),
- [Daniel L. Silbergeld¹⁰](#),
- [Kimberly Smith](#) ORCID: orcid.org/0000-0002-3142-1970¹,
- [Bosiljka Tasic](#) ORCID: orcid.org/0000-0002-6861-4506¹,
- [Rafael Yuste¹²](#),
- [Idan Segev¹³](#),
- [Christiaan P. J. de Kock](#) ORCID: orcid.org/0000-0002-6697-0179⁶,
- [Huibert D. Mansvelder⁶](#),
- [Gabor Tamas](#) ORCID: orcid.org/0000-0002-7905-6001⁵,
- [Hongkui Zeng](#) ORCID: orcid.org/0000-0002-0326-5878¹,
- [Christof Koch](#) ORCID: orcid.org/0000-0001-6482-8067¹ &
- [Ed S. Lein](#) ORCID: orcid.org/0000-0001-9012-6552^{1,10}

Nature volume **601**, page E12 (2022)

- 930 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Cellular neuroscience](#)
- [Molecular neuroscience](#)

The [Original Article](#) was published on 06 October 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03813-8>

Published online 6 October 2021

In the version of this Article initially published, the Acknowledgements statement contained an error. Originally appearing with thanks for support given in part as follows, “R01EY023173 from The National Eye Institute, U01MH105982 from the National Institute of Mental Health and Eunice Kennedy Shriver National Institute of Child Health & Human Development, and R01EY023173 from The National Institute of Allergy and Infectious Disease,” the last number (R01EY023173) was mistakenly added and is not in fact a grant or one provided by the NIAID. The mention has been removed.

The changes have been made to the online version of the Article.

Author information

Author notes

1. These authors contributed equally: Jim Berg, Staci A. Sorensen, Jonathan T. Ting, Jeremy A. Miller

Affiliations

1. Allen Institute for Brain Science, Seattle, WA, USA

Jim Berg, Staci A. Sorensen, Jonathan T. Ting, Jeremy A. Miller, Thomas Chartrand, Anatoly Buchin, Trygve E. Bakken, Agata Budzillo, Nick Dee, Song-Lin Ding, Nathan W. Gouwens, Rebecca D.

Hodge, Brian Kalmbach, Changkyu Lee, Brian R. Lee, Lauren Alfiler, Katherine Baker, Eliza Barkan, Kyla Berry, Darren Bertagnolli, Kris Bickley, Jasmine Bomben, Krissy Brouner, Tamara Casper, Peter Chong, Kirsten Crichton, Rachel Dalley, Rebecca de Frates, Tsega Desta, Samuel Dingman Lee, Florence D’Orazi, Nadezhda Dotson, Tom Egdorf, Rachel Enstrom, Colin Farrell, David Feng, Olivia Fong, Clare Gamlin, Amanda Gary, Alexandra Glandon, Jeff Goldy, Melissa Gorham, Sergey Gratiy, Lucas Graybuck, Hong Gu, Kristen Hadley, Alex M. Henry, DiJon Hill, Chris Hill, Madie Hupp, Tim Jarsky, Sara Kebede, Lisa Kim, Mean-Hwan Kim, Matthew Kroll, Boaz P. Levi, Katherine E. Link, Matthew Mallory, Rusty Mann, Michelle Maxwell, Medea McGraw, Delissa McMillen, Stephanie Mok, Alice Mukora, Lindsay Ng, Kiet Ngo, Philip R. Nicovich, Julie Nyhus, Aaron Oldre, Victoria Omstead, Daniel Park, Hanchuan Peng, Trangthanh Pham, Christina A. Pom, Lydia Potekhina, Ramkumar Rajanbabu, Shea Ransford, David Reid, Christine Rimorin, Augustin Ruiz, David Sandman, Josef Sulc, Susan M. Sunkin, Aaron Szafer, Elliot R. Thomsen, Michael Tieu, Amy Torkelson, Jessica Trinh, Herman Tung, Wayne Wakeman, Katelyn Ward, Grace Williams, Zizhen Yao, Costas Anastassiou, Anton Arkhipov, Amy Bernard, Luke Esposito, Michael J. Hawrylycz, Allan R. Jones, Gabe J. Murphy, Lydia Ng, John W. Phillips, Kimberly Smith, Bosiljka Tasic, Hongkui Zeng, Christof Koch & Ed S. Lein

2. Department of Physiology and Biophysics, University of Washington, Seattle, WA, USA

Jonathan T. Ting, Brian Kalmbach & Gabe J. Murphy

3. Department of Pathology, University of Washington, Seattle, WA, USA

Allison Beller, Lisa Keene, Caitlin Latimer, Desiree Marshall, Erica Melief & C. Dirk Keene

4. byte physics, Berlin, Germany

Thomas Braun

5. MTA-SZTE Research Group for Cortical Microcircuits, Department of Physiology, Anatomy, and Neuroscience, University of Szeged, Szeged, Hungary

Szabina Furdan, Leona Mezei, Norbert Mihut, Gabor Molnar, Gaspar Olah, Attila Ozsvár, Viktor Szemenyei & Gabor Tamas

6. Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research (CNCR), Vrije Universiteit, Amsterdam, The Netherlands

Anna A. Galakhova, Natalia A. Goriounova, Tim S. Heistek, Djai B. Heyer, Eline J. Mertens, Femke Waleboer, René Wilbers, Christiaan P. J. de Kock & Huibert D. Mansvelder

7. Swedish Neuroscience Institute, Seattle, WA, USA

Nathan Hansen, Jae-Geun Yoon, Charles Cobbs & Ryder P. Gwinn

8. Department of Neurosurgery, University of Szeged, Szeged, Hungary

Pal Barzo

9. Cancer Center Amsterdam, Brain Tumor Center, Department of Neurosurgery, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands

Philip C. de Witt Hamer & Sander Idema

10. Department of Neurological Surgery, University of Washington, Seattle, WA, USA

Richard G. Ellenbogen, Manuel Ferreira, Andrew L. Ko, Jeffrey G. Ojemann, Anoop P. Patel, Daniel L. Silbergeld & Ed S. Lein

11. Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY,

USA

Patrick R. Hof

12. NeuroTechnology Center, Columbia University, New York, NY, USA

Rafael Yuste

13. Edmond and Lily Safra Center for Brain Sciences and Department of Neurobiology, The Hebrew University Jerusalem, Jerusalem, Israel

Idan Segev

Corresponding author

Correspondence to [Ed S. Lein](#).

Rights and permissions

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

[Reprints and Permissions](#)

About this article

Cite this article

Berg, J., Sorensen, S.A., Ting, J.T. *et al.* Author Correction: Human neocortical expansion involves glutamatergic neuron diversification. *Nature* **601**, E12 (2022). <https://doi.org/10.1038/s41586-021-04322-4>

- Published: 06 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04322-4>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04322-4>

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

- Publisher Correction
- [Published: 06 January 2022](#)

Publisher Correction: Observation of Stark many-body localization without disorder

- [W. Morong](#) ORCID: [orcid.org/0000-0003-4880-8159¹](https://orcid.org/0000-0003-4880-8159),
- [F. Liu](#) ORCID: [orcid.org/0000-0002-5690-4995¹](https://orcid.org/0000-0002-5690-4995),
- [P. Becker](#) ORCID: [orcid.org/0000-0002-7786-9447¹](https://orcid.org/0000-0002-7786-9447),
- [K. S. Collins](#) ORCID: [orcid.org/0000-0002-6159-8013¹](https://orcid.org/0000-0002-6159-8013),
- [L. Feng¹](#),
- [A. Kyprianidis¹](#),
- [G. Pagano²](#),
- [T. You¹](#),
- [A. V. Gorshkov](#) ORCID: [orcid.org/0000-0003-0509-3421¹](https://orcid.org/0000-0003-0509-3421) &
- [C. Monroe¹](#)

Nature volume **601**, page E13 (2022)

- 766 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Quantum simulation](#)
- [Statistical physics](#)
- [Thermodynamics](#)

- [Ultracold gases](#)

The [Original Article](#) was published on 17 November 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03988-0>
Published online 17 November 2021

In the version of this Article initially published, an error appeared in the third sentence of the Methods section. A stray character “F” appeared to the right of the close bracket at the end of the equation in the sentence now reading “Pseudospins are encoded in the two-clock ground hyperfine states, with $|F = 0, m_F = 0\rangle = |\downarrow_z\rangle$ and $|F = 1, m_F = 0\rangle = |\uparrow_z\rangle$, where F and m_F are the quantum numbers for the total angular momentum and the component along z , respectively.”

The original Article has been corrected online.

Author information

Affiliations

1. Joint Quantum Institute and Joint Center for Quantum Information and Computer Science, University of Maryland and NIST, College Park, MD, USA

W. Morong, F. Liu, P. Becker, K. S. Collins, L. Feng, A. Kyprianidis, T. You, A. V. Gorshkov & C. Monroe

2. Department of Physics and Astronomy, Rice University, Houston, TX, USA

G. Pagano

Corresponding authors

Correspondence to [W. Morong](#) or [F. Liu](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Morong, W., Liu, F., Becker, P. *et al.* Publisher Correction: Observation of Stark many-body localization without disorder. *Nature* **601**, E13 (2022). <https://doi.org/10.1038/s41586-021-04271-y>

- Published: 06 January 2022
- Issue Date: 20 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04271-y>

Share this article

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

[Get shareable link](#)

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

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

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

Collections