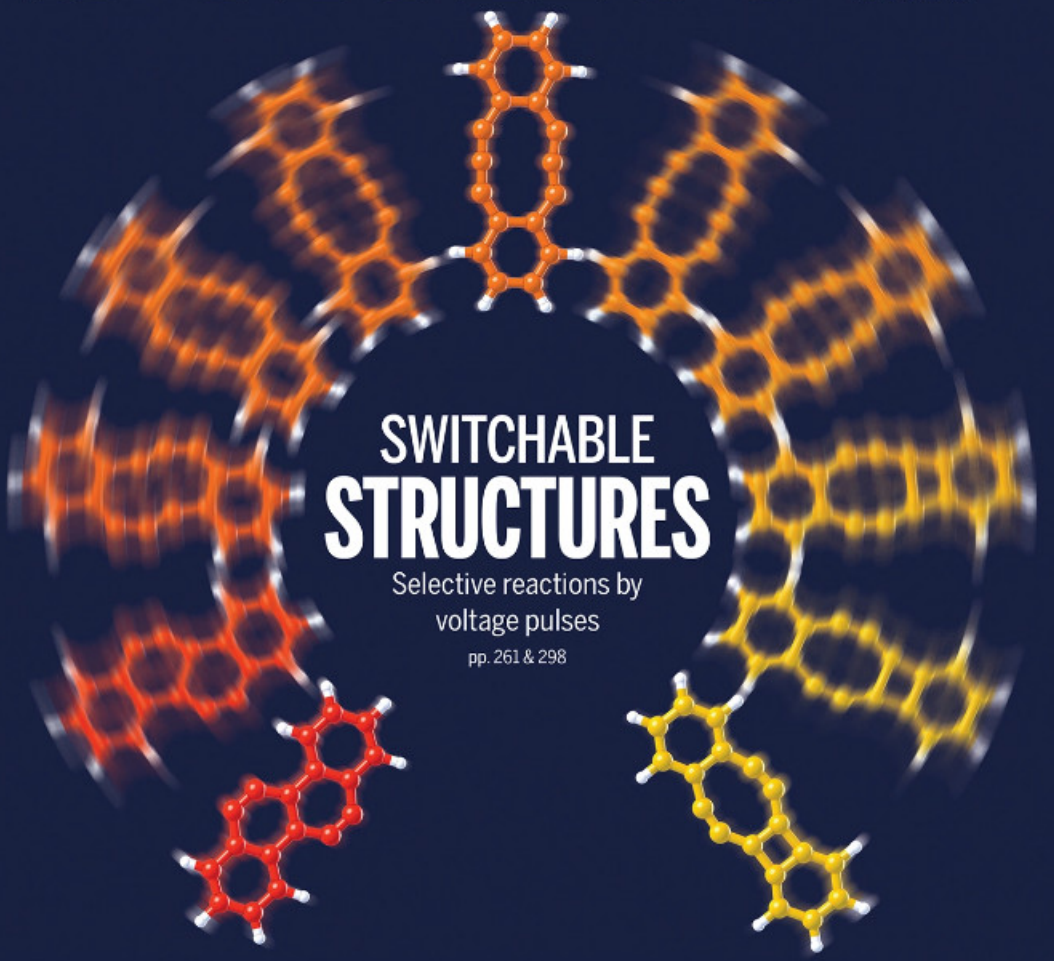
A new bid to tap energy from deep, hot rock p. 252

Tackling the ecological impact of pharmaceutical pollution p. 259

Y chromosome loss linked to cardiac risk pp. 266 & 292

Science

$15
15 JULY 2022
science.org
AAAS



SWITCHABLE STRUCTURES

Selective reactions by voltage pulses

pp. 261 & 298

EDITORIAL

Nailing the nuance on COVID-19

In messaging the public about COVID-19, brevity and simplicity are crucial. But so too is scientifically accurate and nuanced communication that tells people what the latest numbers mean and how they should influence behavior. Nearly 3 years into the pandemic, public health communicators continue to struggle to get this balance right.

It is not an easy task. New scientific information about the virus emerges continually, social media gives voice to unlimited interpretations of new information (some designed to foment confusion), and partisanship can promote ideas designed to support political agendas rather than describe scientific truth. In the face of these difficulties, policy-makers must produce overarching recommendations that are understandable to the public. In doing so, such recommendations occasionally trade accuracy for simplicity. Although sometimes necessary, the message can end up more confusing than clarifying.

Take the current isolation guidelines from the US Centers for Disease Control and Prevention (CDC), under which COVID-19 patients with improved symptoms can end isolation 5 days after the first day of symptoms or a positive viral test, and wear a mask for an additional five. Given that the science clearly shows that most patients continue to be infectious well beyond day 5, a question emerges: Is the CDC's policy anti-science?

Not exactly. Not all patients have access to rapid COVID-19 tests—a gap that makes a testing-based standard problematic. Further, the CDC recognizes that prolonged isolation places hardships on individuals, schools, and businesses. Assuming that patients who are on days 6 to 10 after isolation adhere to the masking recommendation, the threat that they pose to the community likely pales in comparison to that presented by asymptomatic, maskless people who don't know they have COVID-19. Considering this context, the CDC's recommendation may make sense from an overarching policy perspective. However, the guideline is often misinterpreted by the public as meaning that people are incapable of spreading the virus after day 5—which is wrong.

Another example is the change that CDC made earlier this year in categorizing community-level COVID-19 risk, moving to a greater emphasis on hospitalization rates than on community case rates. Under this method, even communities with sky-high case rates are deemed to be at a "medium" risk if hospital admissions and occupancy rates are not excessive. The hospitalization rate is an important metric for health care leaders because as hospitals fill, they may need to take actions to maintain safe operations, such as to cancel elective surgeries. And local policy-makers may see a high hospital COVID-19 census as a trigger for mask mandates or other measures to protect the community.

But for an individual trying to decide whether to eat indoors or wear a mask while shopping, hospitalization rates are nearly meaningless. Instead, the salient question is: What are the odds that a person standing near me in an indoor space has COVID-19? For that, local case rates (even though they are underestimates because of home testing that goes unreported), test positivity rates, and wastewater detection rates provide far more useful information. Yet many regions are now deemed "medium risk" by the CDC despite very high case and test positivity rates—numbers that indicate that transmission risk to individuals is quite high.

> "The need to clarify messages and their meanings will be an enduring lesson of the pandemic."

For CDC and other health officials, more thought must be put into the intent of providing the public with information about particular rates and numbers. For example, rather than implying that most people are not contagious after day 5 of infection, it would be preferable for officials to explain that for a patient with improving symptoms, the chance of spreading infection to others is low (but not zero) after day 5, which is why further isolation is not required. However, to be safe to others, individuals exiting isolation should wear a mask until day 10.

And rather than implying that low hospitalization rates mean that individuals are not at high risk for COVID-19 transmission, it might be best to explain that although COVID-19–related hospital occupancy is relatively low (thus hospitals are not under stress), the case and test positivity rates are high, and so the opportunity for transmission is also high. Thus, mask wearing in crowded indoor spaces is strongly recommended.

Whereas breathtaking advances in vaccine science have saved millions of lives, the communication of complex and nuanced information has lagged. The need to clarify messages and their meanings will be an enduring lesson of the pandemic.

–Robert M. Wachter

PHOTO: CHRISTOPHER MICHEL

Robert M. Wachter is a professor and chair of the Department of Medicine at the University of California, San Francisco, CA, USA. robert.wachter@ucsf.edu

10.1126/science.add8599

NEWS

> “Let’s hope that some fresh political faces can move the situation forward.”
>
> **Research policy scholar Graeme Reid**, in *Times Higher Education*, on prospects that Prime Minister Boris Johnson’s resignation will help U.K. researchers access the European Union’s flagship science funding program.

IN BRIEF

Edited by **Shraddha Chakradhar** *and* **Jeffrey Brainard**

PUBLIC HEALTH

Beijing flips on vaccine mandate

City authorities in Beijing announced China’s first large-scale COVID-19 vaccine mandate on 6 July, only to scrap the plan a day later after social media posts questioned the measure’s legality and the vaccines’ effectiveness. (China’s censors have allowed citizens to vent complaints about the handling of COVID-19 outbreaks as long as they don’t question the government’s overall “zero COVID” strategy.) The mandate called for those entering libraries, museums, movie theaters, gyms, and other public facilities to show proof of COVID-19 vaccination, a policy enacted by many countries during the pandemic. Instead, Beijing reverted to requiring proof of a negative COVID-19 test taken in the previous 72 hours and a temperature check to visit facilities that attract crowds. Roughly 90% of China’s population has gotten vaccinated voluntarily, but only half of those older than 80 were vaccinated as of March, the last time national details were released. Many unvaccinated seniors fear side effects of inoculation, making the national government reluctant to relax its zero COVID strategy, which relies on contact tracing, mass testing, and lockdowns.

EU’s ‘sustainable’ power knocked

ENERGY POLICY | EU lawmakers last week voted 328 to 278 to define natural gas and nuclear power as “sustainable” energy sources, drawing strong objections from environmental groups. The designation, proposed earlier this year by the European Commission, is aimed at easing private investment in energy sources that some see as less polluting than coal and oil. But critics say the policy ignores scientific evidence, will only complicate efforts to curb climate change, and could end up diverting funding from solar and wind projects to relatively expensive nuclear plants. Opponents are urging EU member nations to stop the policy from going into effect next year and say they will go to court if that strategy fails. In 2020, natural gas provided about 24% of the European Union’s total energy, whereas nuclear provided 13%.

Embattled spider biologist resigns

RESEARCH INTEGRITY | A behavioral ecologist under fire for more than 2 years for data irregularities or possible fabrication in dozens of publications resigned this week from a prestigious, tenured position at McMaster University. Jonathan Pruitt’s work on social behavior in spiders had earned international acclaim, and their willingness to share data drew many eager collaborators. In a statement, McMaster confirmed it has reached a “confidential” settlement with Pruitt, but said it is not ready to release the findings of its probe into possible research misconduct. Pruitt told *Science* they are not yet able to speak publicly about any allegations. Some scientists who co-authored papers with Pruitt criticized McMaster for acting too slowly on the misconduct concerns, first made public in 2020, and urged the university to unveil its findings now.

Marburg virus hits Ghana

INFECTIOUS DISEASES | Two people infected with the Marburg virus have died in Ghana, the first ever detected cases of the deadly hemorrhagic fever in the country, health authorities report. Tissue samples collected from the people, in the Ashanti region in the south, have been sent to the Pasteur Institute in Senegal for confirmation, the World Health Organization announced last week. Marburg is closely related to the Ebola virus, but outbreaks are rarer and typically smaller; the biggest one, in Angola in 2004–05, ended after 252 cases. There are no approved vaccines or treatments. It is only the second time a Marburg infection has been detected in West Africa. A single case was confirmed in Guinea in 2021, but no further cases were found and the outbreak was declared over after 5 weeks.

Dino’s puny arms resemble *T. rex*’s

PALEONTOLOGY | For dinosaurs, tiny arms may have been the price of a giant, carnivorous head, according to a study of a new species. In Argentina’s Patagonian Desert, paleontologists discovered a half-complete, 11-meter-long skeleton that’s a *Tyrannosaurus rex* doppelgänger, with stubby arms and a cartoonishly big cranium, but is only distantly related to the tyrannosaurids. The team named their “lucky strike” discovery *Meraxes gigas* after a Targaryen dragon from *Game of Thrones*. Because researchers determined the head and arms

This reconstruction of *Meraxes gigas* may look familiar, but the newly discovered dinosaur was only distantly related to *Tyrannosaurus rex*.

ILLUSTRATION: CARLOS PAPOLIO

went together, they could compare the fossil with other known members of its family, the carcharodontosaurids. They found that over time, the dinosaurs' heads became larger and forearms shrank. Two other families of giant dinosaurs—tyrannosaurids and abelisaurids—independently exhibit similar trends. The big crania supported big jaws that helped the carnivores capture large prey, researchers say, whereas the forelimbs likely shrank to keep the bipedal creatures in balance or as a developmental compensation for a larger skull, not for any adaptive advantage of their own, the team reported in the 7 July issue of *Current Biology*.

U.S. aims to tame quantum threat

CRYPTOGRAPHY | The U.S. National Institute of Standards and Technology (NIST) announced last week it will add to its standards four encryption algorithms that are supposed to be immune to hacking by a quantum computer. One algorithm is for general encryption and three are for digital signatures. The standards are part of a 6-year push to safeguard internet communications from quantum computers, which manipulate bits of information dubbed qubits that can be set to 0 or 1 simultaneously. Scientists expect a large quantum computer will be able to crack the current "public key" encryption algorithms. NIST will consider four other algorithms for general encryption for inclusion in its standards, which the agency expects to finalize in 2 years.

Teaching program boosts voting

POLITICAL SCIENCE | A U.S. program that trains recent college graduates to teach in high-needs schools also makes them more likely to vote, a study has found. Political scientists Cecilia Mo of the University of California, Berkeley, and John Holbein of the University of Virginia examined the impact of Teach for America (TFA) on political engagement by applying a statistical tool to create a group of trainees and a control group of rejected applicants; the two groups' demographic characteristics differ minimally. The TFA alumni vote at rates up to nine percentage points higher, the researchers report in this week's issue of the *Proceedings of the National Academy of Sciences*, in the first such study of a U.S. national service program. Serving in TFA was 14 times more effective in promoting voting by eligible young people than get-out-the-vote campaigns or other direct appeals. It's not clear why TFA improves the dismally low voting record of those under age 30, but the authors say possible reasons include prolonged exposure to social inequalities.

Indigenous groups helped protect the Amazon's pirarucu, one of the largest freshwater fishes, from commercial overharvest.

PHOTO: PETE OXFORD/MINDEN PICTURES

CONSERVATION

United Nations calls use of wild species unsustainable

One in five people depends directly on wild animals, plants, and other organisms for their food or livelihood, but many of them consume the resources unsustainably, according to the first global evaluation of these practices. Billions of people are tapping 55,000 wild species, and some uses are intensifying as the population grows, according to a report issued on 8 July by the United Nations's Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. The body spent 4 years gathering input from researchers, policymakers, and Indigenous groups for its *Assessment Report on the Sustainable Use of Wild Species*. Fixes include stepping up regulation of supply chains and curtailing illegal trade of heavily consumed organisms such as fish or the use of timber for cooking; one in three people globally relies on wood this way, the report found. It also emphasizes the need for policies to be inclusive, equitable, and tailored to ensure Indigenous people continue to survive and manage their resources.

IN DEPTH

The James Webb Space Telescope's new view of the Carina nebula, a stellar nursery full of gas and dust, improves on images from the Hubble Space Telescope.

ASTRONOMY

Webb telescope wows with first images

Pictures showcase start of science campaign for NASA's largest space telescope

By **Daniel Clery**

At long last, eager astronomers—and the public—get to view snapshots of the universe provided by the James Webb Space Telescope, the largest, most complex, and most expensive space telescope ever built. On 11 July, in a sneak preview, the White House released the first public picture, a deep image bristling with thousands of distant galaxies. The next day, NASA released four more that demonstrate Webb's ability to scrutinize gas clouds and planetary systems closer to home.

The images—a product of Webb's majestic, gold-plated mirror—are not only sharper than those of the Hubble Space Telescope, but are also essentially different, capturing the longer, infrared wavelengths that are important for many branches of astronomy. For Webb's engineers and designers, who have endured years of delay and a nerve-wracking 6 months of launch, deployment, and commissioning, the relief is palpable. "I'm feeling a mix of excitement and emotion. [Webb] has really delivered, and there will be a rush of new discoveries," says Brant Robertson of the University of California (UC), Santa Cruz, who helped develop Webb's near-infrared camera.

For most astronomers, the wait isn't quite over yet, as their scheduled observations may not happen for weeks or months. Exoplanet researcher Laura Kreidberg of the Max Planck Institute for Astronomy (MPIA) is one of the lucky ones. As part of a large collaboration doing "early release science" with Webb, she'll get to see a first tranche of data in 2 days. Nevertheless, "It's a long wait from Tuesday to Thursday," she says. But come Thursday afternoon when the data start to pour in, "I'll make a big pot of coffee, put out some snacks, and we'll sit and look at it."

Conceived in the 1990s and built with European and Canadian contributions, Webb at times seemed like a cursed mission, buffeted by cost overruns, schedule slips, and technical hitches. Beset by criticism, "It was hard to keep the team at the edge of excellence," NASA science chief Thomas Zurbuchen said at a recent briefing. But after a flawless launch on 25 December 2021, those trials are almost forgotten. "I don't think any of us expected how well it went," Zurbuchen says.

Webb's view of the Southern Ring nebula reveals a glowing shell of gas blown off by a giant, dying star.

The accuracy of the launch, on a European Ariane 5 rocket, meant Webb didn't have to use much fuel for course corrections and now has more to keep itself in a gravitational pocket on the opposite side of Earth from the Sun, about 1.5 million kilometers away. Planned as a 10-year mission, Webb is now expected to last at least twice as long.

The launch of the $10 billion instrument did not end the tension. To unfurl its giant sunshield, swing six of the 18 segments in the 6.5-meter-wide mirror into position, and extend the secondary mirror on its booms, engineers had to navigate some 300 steps, any one of which could have doomed the mission. "Every day the risk level has gone down and my ability to sleep has increased," says Charlie Atkinson, the project's chief engineer at Northrop Grumman, NASA's prime contractor for the mission.

Tiny motors adjusted the position, tilt, and curvature of the mirror segments by fractions of a hair's breadth until they could together focus on targets as a single mirror. Operators then had to check out Webb's four instruments, a mix of cameras and spectrographs, which split incoming light into its component wavelengths. "We have an observatory in excellent shape, that meets or exceeds expectations," Bill Ochs, Webb's project man-

IMAGES: NASA; ESA; CSA; STSCI

ager at NASA's Goddard Space Flight Center, announced on 12 July.

Not everything went according to plan. Some computer glitches required a few hours' pause. The mirror has been dinged by a handful of micrometeorites—that was expected, but one was larger than models predicted and operators are working to mitigate its impact. "There have been some issues," Ochs says. "But when you have a good team, you can get through it."

By 21 June, Webb was taking data for the 286 teams of scientists who had been allotted time on the telescope in its first year, known as cycle one. "We've got some amazing science in the can," Zurbuchen says. Researchers have reported seeing that some of their scheduled observations have taken place, but NASA sat on the data until this week, when commissioning officially ends. Now, Atkinson says, "We hand over the keys. It's ready. Go do science."

The pictures and spectra released today are the result of a yearlong selection process that Zurbuchen calls "bottom-up." NASA wanted to show off the telescope's capabilities and offer a taste of the different fields of astronomy it will transform. So managers asked the instrument and science teams for a range of targets that would show "the best of this thing," Zurbuchen says. Some 70 were suggested and these were whittled down by a committee to the five released so far.

A picture of galaxy cluster SMACS 0723 showcases Webb's ability to peer into the most distant corners of the universe and hence the furthest back in time. It shows a field crowded with thousands of galaxies, some with shapes distorted by the intense gravitational field of the cluster in the foreground. A spectrum from a 13.1-billion-year-old galaxy in the image shows it contained oxygen, hydrogen, and neon—the most distant galaxy for which the constituents are known. Dominika Wylezalek of Heidelberg University calls the image "mind-blowing." "The level of detail is really breathtaking."

Hubble saw further back than ever imagined, spying a galaxy that existed just 400 million years after the big bang—3% of the universe's current age. But Webb will see many more galaxies that are even younger, not just because of its larger mirror, but also because of its sensitivity to infrared light. Photons emitted by the earliest stars are stretched in their journey by the expansion of the universe, pushing them into infrared wavelengths that Hubble cannot see. Previously, says MPIA's Sarah Bosman, "We could only see the very bright galaxies, and the biggest. With Webb, we'll see the whole array."

Galaxy surveys will shed light on the early history of these agglomerations: when they started to form stars and how quickly they organized into the disklike spirals such as the Milky Way. "Webb is capable of filling in the gaps," Robertson says. Webb will also help figure out what ionized the neutral hydrogen gas that filled the universe before the stars turned on. By the time of the universe's billionth birthday, that hydrogen was ionized. Astronomers think this was mostly the work of high-energy ultraviolet photons from the first generation of stars, which were huge and bright and made solely of this primordial hydrogen. But Hubble hasn't seen enough early galaxies to account for the needed photons. "Webb is the only facility that can see" this early era, Robertson says.

Webb's view of Stephan's Quintet, four merging galaxies with a fifth in the foreground, reveals glowing gas heated by the shock of merger.

Another one of the released pictures is of the well-known Stephan's Quintet, a cluster of four interacting galaxies 290 million light-years away that sits behind a fifth galaxy in the foreground. Seen with Webb, it is possible to discern glowing gas and dust heated by two of the merging galaxies and areas of active star formation generated by that turmoil. "It is this sort of interaction that drives the evolution of galaxies," says Giovanna Giardino of the European Space Agency.

Alice Shapley of UC Los Angeles will use Webb's near-infrared spectrograph (NIRSpec) to tease apart the light from galaxies to find out how hot they are, how they're moving, and what they're made of. She's interested in faint emission lines from interstellar oxygen gas, which show up as spikes in spectra. The oxygen is created in massive stars and scattered when they die. "It tells you how many stars are made," she says, and is a good marker for the flows of gas in and out of a galaxy.

The information will help researchers understand why some galaxies are prolific creators of stars, whereas others are subdued or even dead. Earth's atmosphere clouds the view of these oxygen lines for most ground-based spectrographs. NIRSpec uses a mask covered with tiny shutters to gather light and generate spectra from dozens of galaxies simultaneously, which should increase the number of galaxies with known oxygen lines by an order of magnitude. "Webb allows us to go so much further," Shapley says. "It's going to be amazing."

Misty Bentz of Georgia State University will test a different NIRSpec capability: taking a spectroscopic "image" in which every pixel has its own separate spectrum. Her project involves staring for 9 hours at a single galaxy, the nearby NGC 4151—dubbed the Eye of Sauron because of the eerie glow of the supermassive black hole at its heart, which shines brightly as it heats up gas being drawn into its maw. Bentz will look for subtle changes to the spectrum of the swirling gas that reveal the black hole's mass, showcasing NIRSpec's ability to take faint spectra of the gas while blocking the bright area around the black hole itself.

Webb will also study targets within the Milky Way—objects like the Carina nebula, one of today's released pictures. Webb's image of the vast stellar nursery, 7600 light-years from Earth, contains hundreds of newly born stars that have never been seen before, as well as swirling dust and gas buffeted by stellar wind. Hubble also imaged this local landmark, but Webb's image has "so much more detail," says NASA Deputy Project Scientist Amber Straughn.

Within such stellar nurseries, dense clumps of gas gradually collapse to form stars. Melissa McClure of the Leiden Observatory will use Webb to peer into such clouds to see whether they are factories for complex molecules that could give life a head start even before star and planet formation begins. Hundreds of different molecules have been detected as gases, but gases don't react efficiently unless frozen onto solid interstellar grains of dust. The grains are like singles bars, McClure jokes, where atoms and simple molecules go to join up. So far, only methanol ice has been detected in space, but McClure expects Webb will routinely find ices of molecules such as methane and ammonia. The real prize would be finding complex carbon-based molecules with more than six atoms, such

IMAGE: NASA; ESA; CSA; STScI

as ethanol or acetaldehyde. "A conclusive detection would be really awesome," McClure says.

Among NASA's offerings this week, the most tantalizing taste of what's to come is the spectrum of an exoplanet. It's from a giant planet known as WASP-96b orbiting close to a star 1150 light-years from Earth. During regular passes in front of its home star, some starlight is absorbed by gases in WASP-96b's atmosphere, leaving telltale dips in the star's spectrum. Models suggest carbon monoxide, carbon dioxide, and methane may be present, but Hubble and ground-based telescopes couldn't see them. The first spectrum of WASP-96b shows clear signs of water vapor and features that indicate the presence of clouds and vapor haze. Astronomers will have to wait for future transits to see what else is there. "I'm really excited to see if the predictions bear out," Kreidberg says.

Using the Hubble and Spitzer space telescopes, researchers have spied water and sodium in exoplanet atmospheres. "It's been like reading a poem and only seeing every third word," Kreidberg says. Webb is expected to find a wealth of molecules, some of which can hint at a planet's potential habitability: carbon monoxide and dioxide, methane, ammonia, phosphine, and more. They will be able to probe the atmospheres of every kind of planet from hot Jupiters, through mini-Neptunes, to rocky planets like Earth. Hundreds of exoplanet researchers are preparing feverishly for the expected data deluge from other transiting exoplanets, organizing hackathons and data challenges, and devising data processing pipelines. "We've done a huge amount of work to make sure we're ready," Kreidberg says. "There's no calm before the storm."

A fraction of Webb's time will go to objects in the Solar System, but for targets so close and so large—from Webb's point of view—saturating the instruments is a real problem, says Imke de Pater of UC Berkeley. She will study Jupiter's thin ring, which may still be rippling in the wake of the 1994 impact of Comet Shoemaker-Levy 9. Imaging the faint ring next to the much brighter planet will be a challenge, but she hopes to discover more ripples, which could signal more recent, unseen comet impacts. De Pater's fondest wish would be detecting new moonlets of just a few hundred meters diameter or less. "That would be a dream," she says.

For many more, the dream has already come true—an effort that took 30 years and some 20,000 people around the world to bear fruit. "Beyond the science, [Webb] is rekindling a sense of beauty and wonder about the universe that inspired me to become an astronomer," Robertson says. "I couldn't be happier." ■

U.S. RESEARCH FUNDING

Bill would bar NIH and CDC from funding lab research in China

House of Representatives measure catalyzed in part by suspicions that Wuhan lab leak led to pandemic

By **Jocelyn Kaiser**

A proposal moving through Congress to bar the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC) from funding research laboratories in China is sparking concern among scientists. If signed into law, the measure could cut off millions of dollars of U.S. funds flowing to collaborative research projects in several areas, including HIV/AIDS, cancer, mental health, and flu surveillance.

The proposed ban, part of a 2023 spending bill approved by the U.S. House of Representatives Committee on Appropriations on 30 June, grew out of suspicions among some lawmakers, so far unsupported by evidence, that the Wuhan Institute of Virology (WIV) in China released the coronavirus that started the current pandemic, as well as objections to other potentially risky biomedical experiments involving animals. Specifically, the measure would bar the Department of Health and Human Services (the parent agency of NIH and CDC) from funding WIV or "any other laboratory" in China, Russia, or any country the U.S. government has designated a foreign adversary, a list that currently includes Iran and North Korea.

$5.6 million
Value of NIH grants to China this year

The measure's sponsor, Representative Chris Stewart (R–UT), says the ban is aimed at ensuring the United States does not fund "dangerous research" in "uncontrolled environments" overseas. Backing the measure is the White Coat Waste Project, an animal rights group that 2 years ago publicized NIH's funding of WIV.

Some scientific organizations are concerned by the proposal's expansive scope. "It seems a bit extreme," says Eva Maciejewski, spokesperson for the Foundation for Biomedical Research, which advocates for animal research. "In theory it's good to have oversight over biosafety and animal welfare, but in practice there may be better ways than blocking all NIH funding to foreign countries."

An NIH spokesperson said the agency does not comment on pending legislation. But Gerald Keusch of Boston University, a former director of the NIH Fogarty International Center, believes "most of the senior leadership [of NIH's 27 institutes] will be deeply concerned to have Congress interfering in the review and awarding of grants." (It is unusual for lawmakers to adopt such countrywide bans on research funding.)

The ban's potential impact isn't clear. WIV is largely funded by the Chinese government, and researchers there have received no U.S. funding since NIH, citing compliance issues, suspended a small subcontract for studying bat coronaviruses in July 2021. But NIH supports other research in China, with grants totaling $8.9 million in 2021 and $5.6 million this year, according to federal databases.

Projects that do not involve laboratory work—such as a long-running NIH-funded survey on health and retirement in China—could be spared. But many others would likely be vulnerable, including three projects headed by Chinese investigators studying influenza and the mosquito-borne diseases dengue and malaria, and dozens of subawards to Chinese groups participating in clinical trials of drugs, studies of the health effects of heavy metals, and neuroscience research. The U.S. leader of one clinical trial in Shanghai—who asked for anonymity—said his Chinese partner is a former trainee and "close collaborator," and it would not be possible to recruit enough patients at a single site in the United States.

The ban would likely have a smaller impact on research in Russia. NIH and CDC appear to have just two active grants there, and they may already be subject to a recent White House guidance winding down U.S. research funding to Russia because of its war against Ukraine. There are no NIH or CDC grants to researchers in Iran or North Korea.

To become law, the ban would need to survive negotiation of a final bill with the Senate. Some research groups hope lawmakers will remove the provision before any bill goes to President Joe Biden for final approval, likely late this year. ■

as ethanol or acetaldehyde. "A conclusive detection would be really awesome," McClure says.

Among NASA's offerings this week, the most tantalizing taste of what's to come is the spectrum of an exoplanet. It's from a giant planet known as WASP-96b orbiting close to a star 1150 light-years from Earth. During regular passes in front of its home star, some starlight is absorbed by gases in WASP-96b's atmosphere, leaving telltale dips in the star's spectrum. Models suggest carbon monoxide, carbon dioxide, and methane may be present, but Hubble and ground-based telescopes couldn't see them. The first spectrum of WASP-96b shows clear signs of water vapor and features that indicate the presence of clouds and vapor haze. Astronomers will have to wait for future transits to see what else is there. "I'm really excited to see if the predictions bear out," Kreidberg says.

Using the Hubble and Spitzer space telescopes, researchers have spied water and sodium in exoplanet atmospheres. "It's been like reading a poem and only seeing every third word," Kreidberg says. Webb is expected to find a wealth of molecules, some of which can hint at a planet's potential habitability: carbon monoxide and dioxide, methane, ammonia, phosphine, and more. They will be able to probe the atmospheres of every kind of planet from hot Jupiters, through mini-Neptunes, to rocky planets like Earth. Hundreds of exoplanet researchers are preparing feverishly for the expected data deluge from other transiting exoplanets, organizing hackathons and data challenges, and devising data processing pipelines. "We've done a huge amount of work to make sure we're ready," Kreidberg says. "There's no calm before the storm."

A fraction of Webb's time will go to objects in the Solar System, but for targets so close and so large—from Webb's point of view—saturating the instruments is a real problem, says Imke de Pater of UC Berkeley. She will study Jupiter's thin ring, which may still be rippling in the wake of the 1994 impact of Comet Shoemaker-Levy 9. Imaging the faint ring next to the much brighter planet will be a challenge, but she hopes to discover more ripples, which could signal more recent, unseen comet impacts. De Pater's fondest wish would be detecting new moonlets of just a few hundred meters diameter or less. "That would be a dream," she says.

For many more, the dream has already come true—an effort that took 30 years and some 20,000 people around the world to bear fruit. "Beyond the science, [Webb] is rekindling a sense of beauty and wonder about the universe that inspired me to become an astronomer," Robertson says. "I couldn't be happier." ■

U.S. RESEARCH FUNDING

Bill would bar NIH and CDC from funding lab research in China

House of Representatives measure catalyzed in part by suspicions that Wuhan lab leak led to pandemic

By **Jocelyn Kaiser**

A proposal moving through Congress to bar the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC) from funding research laboratories in China is sparking concern among scientists. If signed into law, the measure could cut off millions of dollars of U.S. funds flowing to collaborative research projects in several areas, including HIV/AIDS, cancer, mental health, and flu surveillance.

The proposed ban, part of a 2023 spending bill approved by the U.S. House of Representatives Committee on Appropriations on 30 June, grew out of suspicions among some lawmakers, so far unsupported by evidence, that the Wuhan Institute of Virology (WIV) in China released the coronavirus that started the current pandemic, as well as objections to other potentially risky biomedical experiments involving animals. Specifically, the measure would bar the Department of Health and Human Services (the parent agency of NIH and CDC) from funding WIV or "any other laboratory" in China, Russia, or any country the U.S. government has designated a foreign adversary, a list that currently includes Iran and North Korea.

$5.6 million
Value of NIH grants to China this year

The measure's sponsor, Representative Chris Stewart (R–UT), says the ban is aimed at ensuring the United States does not fund "dangerous research" in "uncontrolled environments" overseas. Backing the measure is the White Coat Waste Project, an animal rights group that 2 years ago publicized NIH's funding of WIV.

Some scientific organizations are concerned by the proposal's expansive scope. "It seems a bit extreme," says Eva Maciejewski, spokesperson for the Foundation for Biomedical Research, which advocates for animal research. "In theory it's good to have oversight over biosafety and animal welfare, but in practice there may be better ways than blocking all NIH funding to foreign countries."

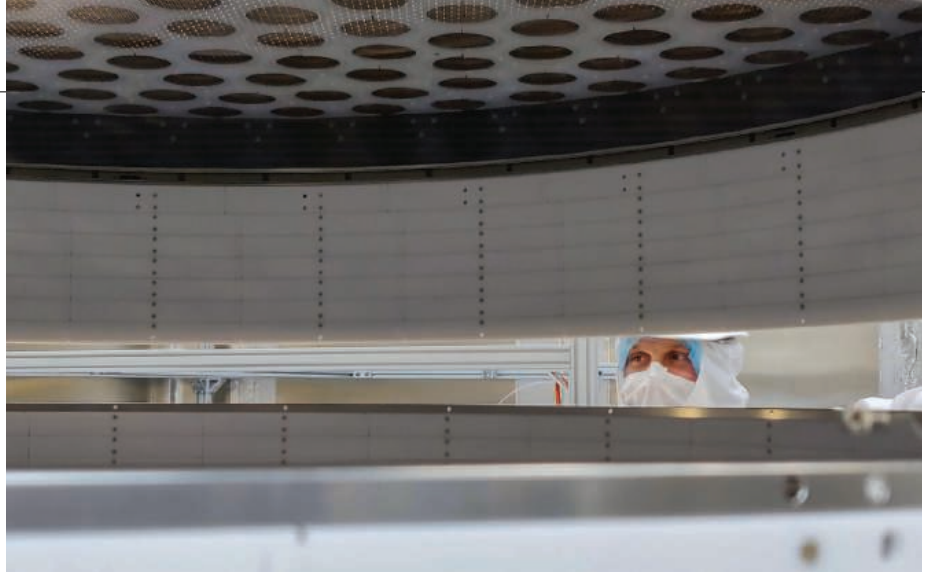
An NIH spokesperson said the agency does not comment on pending legislation. But Gerald Keusch of Boston University, a former director of the NIH Fogarty International Center, believes "most of the senior leadership [of NIH's 27 institutes] will be deeply concerned to have Congress interfering in the review and awarding of grants." (It is unusual for lawmakers to adopt such countrywide bans on research funding.)

The ban's potential impact isn't clear. WIV is largely funded by the Chinese government, and researchers there have received no U.S. funding since NIH, citing compliance issues, suspended a small subcontract for studying bat coronaviruses in July 2021. But NIH supports other research in China, with grants totaling $8.9 million in 2021 and $5.6 million this year, according to federal databases.

Projects that do not involve laboratory work—such as a long-running NIH-funded survey on health and retirement in China—could be spared. But many others would likely be vulnerable, including three projects headed by Chinese investigators studying influenza and the mosquito-borne diseases dengue and malaria, and dozens of subawards to Chinese groups participating in clinical trials of drugs, studies of the health effects of heavy metals, and neuroscience research. The U.S. leader of one clinical trial in Shanghai—who asked for anonymity—said his Chinese partner is a former trainee and "close collaborator," and it would not be possible to recruit enough patients at a single site in the United States.

The ban would likely have a smaller impact on research in Russia. NIH and CDC appear to have just two active grants there, and they may already be subject to a recent White House guidance winding down U.S. research funding to Russia because of its war against Ukraine. There are no NIH or CDC grants to researchers in Iran or North Korea.

To become law, the ban would need to survive negotiation of a final bill with the Senate. Some research groups hope lawmakers will remove the provision before any bill goes to President Joe Biden for final approval, likely late this year. ■



A researcher helps assemble the inner chamber of the LUX-ZEPLIN detector, which now holds 7 tons of frigid liquid xenon.

PHOTO: LAWRENCE BERKELEY NATIONAL LABORATORY

PHYSICS

Hunt for a long-sought dark matter particle nears a climax

Three huge detectors are racing to detect WIMPs, long the favored candidate for universe's missing matter

By **Adrian Cho**

No one was surprised last week when researchers with a massive new dark matter detector called LUX-ZEPLIN (LZ) reported finding nothing. After all, LZ started hunting its quarry, hypothetical weakly interacting massive particles (WIMPs), a few months ago and has collected just a few percent of its expected data. Still, the announcement piqued the interest of physicists. LZ, a U.S. detector, is turning on at the same time as similar detectors in Italy and China. They likely mark the final or nearly final phase in the decadeslong quest to spot these particles, a potential solution to one of physics' greatest mysteries.

The three-way competition to spot WIMPs will push all the groups to work harder, says Dan Hooper, a theorist at Fermi National Accelerator Laboratory, who was not involved in the work. "The physicist-capitalist in me thinks this is a good thing," Hooper says.

Dark matter is thought to account for 85% of all matter. Astronomical observations show, for example, that the stars in a typical galaxy swirl so fast that their collective gravity isn't strong enough to keep them from flying into space. So physicists assume the extra gravity needed to rein in the stars comes from invisible dark matter—presumably consisting of some new fundamental particle.

Since the 1980s, many physicists have favored WIMPs, which would interact with ordinary matter just through gravity and the weak nuclear force. WIMPs would have emerged naturally after the big bang, and enough of them should linger to account for dark matter, provided they are about 100 times as massive as a proton. They would drift unobtrusively through the Galaxy and even our own bodies, but occasionally one ought to crash into an atomic nucleus. So, physicists stalk WIMPs by looking for recoiling nuclei in detectors deep underground, where they're shielded from radiation that can produce extraneous signal.

For 20 years, scientists have developed ever-bigger detectors consisting of tanks of liquid xenon lined on top and bottom with light-detecting phototubes, which detect the flash of light produced by a recoiling nucleus. Simultaneously, an electric field tugs electrons liberated by the speeding nucleus out of the liquid, producing a second flash. Comparing the sizes and times of the flashes, researchers can distinguish recoiling nuclei from, say, recoiling electrons, which can be generated by gamma rays hitting the detector.

Filled with 7 tons of liquid xenon and hunkering 1480 meters down in the Sanford Underground Research Facility in Lead, South Dakota, LZ is the biggest liquid xenon detector yet. In its first 65 days of data taking, it detected 335 recoil events, Hugh Lippincott, a physicist at the University of California (UC), Santa Barbara, and spokesperson for the 287-member LZ team, reported on 7 July in an online seminar. But roughly that many background events are expected from inevitable traces of radioactive radon in the liquid and other sources, Lippincott reported, so there's no evidence of recoils coming from WIMPs.

That null result puts the strongest limits yet on how strongly WIMPs with masses between about 10 and 10,000 times that of a proton could interact with ordinary matter. LZ's new limits edge past those published in 2021 by a team using PandaX-4T, a 3.7-ton liquid xenon detector in the China Jinping Underground Laboratory. A third detector, XENONnT, has a 5.9-ton central chamber and has been taking data in Italy's subterranean Gran Sasso National Laboratory. Initial results are due later this year, says Rafael Lang, a physicist and XENON team member at Purdue University.

Some of the enthusiasm for WIMPs as dark matter candidates has waned in recent years, not only because searches have come up empty so far, but also because the world's biggest atom smasher, Europe's Large Hadron Collider, has yet to blast out anything that looks like a WIMP. However, physicists are only now starting to probe the heart of the possible ranges of mass and interaction strength for WIMPs, Lang says. "Half of the [possibilities] that you were excited about a decade or two ago are still alive and well."

Largely by chance, all three new detectors have come to life at roughly same time. "Five years ago, you couldn't have predicted them to all come on within months of each other," Lippincott says. Having three different experiments running simultaneously provides an advantage should any of them see a signal, says Katherine Freese, a theoretical astrophysicist at the University of Michigan, Ann Arbor. "You always need to reproduce the results with a different experiment."

The current detectors should run for 3 to 5 years. In case they come up empty, WIMP hunters are already sketching out the ultimate liquid xenon detector, an 80-ton giant. In June, the LZ and XENON teams met to begin planning such a device. "We aim to combine the best aspects of the two [current detectors]," says XENON member Laura Baudis, a physicist at the University of Zürich.

A detector that large would probably mark the end of the line. It would be sensitive enough to begin to detect a "fog" of particles called neutrinos from the Sun and the atmosphere, which it could not distinguish from WIMPs, notes Jonathan Feng, a theorist at UC Irvine. An even bigger liquid xenon detector would make little sense. Still, Feng says, "There's no good reason to stop before you get to the neutrino fog." ■

The Sudd, a swamp in South Sudan that is Africa's largest, has grown as a methane hot spot.

CLIMATE CHANGE

Ominous feedback loop may be accelerating methane emissions

Tropical wetlands, getting wetter with global warming, emerge as hot spots for heat-trapping gas

By **Paul Voosen**

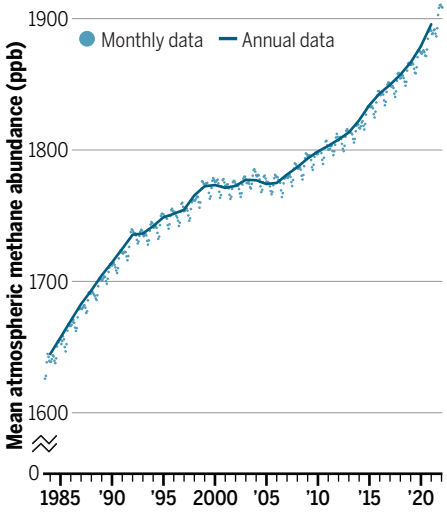
If carbon dioxide is an oven steadily roasting our planet, methane is a blast from the broiler: a more potent but shorter lived greenhouse gas that's responsible for roughly one-third of the 1.2°C of warming since preindustrial times. Atmospheric methane levels have risen nearly 7% since 2006, and the past 2 years saw the biggest jumps yet, even though the pandemic slowed oil and gas production, presumably reducing methane leaks. Now, researchers are homing in on the source of the mysterious surge. Two new preprints trace it to microbes in tropical wetlands. Ominously, climate change itself might be fueling the trend by driving increased rain over the regions.

If so, the wetlands emissions could end up being a runaway process beyond human control, although the magnitude of the feedback loop is uncertain. "We will have handed over a bit more control of Earth's climate to microorganisms," says Paul Palmer, an atmospheric chemist at the University of Edinburgh and co-author of one of the studies, posted late last month for review at *Atmospheric Chemistry and Physics*.

Most climate scientists already agreed that the post-2006 methane spike has largely not come from fossil fuel production. That's because atmospheric methane has become ever more enriched in carbon-12, the lighter isotope of carbon, reversing what had been a multicentury trend, says Xin Lan, a carbon cycle scientist at the Earth System Research Laboratories (ESRL) of the National Oceanic and Atmospheric Administration. "This is a very significant signal," she says. It points to microbes as the source because they favor reactions that use light carbon, giving the methane they produce a distinctive light signature.

Methane on the move

After a brief plateau in the early 2000s, atmospheric methane levels have grown rapidly, causing about one-third of modern global warming. Tropical wetlands may be fueling an acceleration seen after 2019.



Yet the isotopic signal cannot distinguish between microbes in a swamp, a landfill, or a cow's gut. "A cow is a walking wetland," says Euan Nisbet, an atmospheric chemist at Royal Holloway, University of London. Most researchers think a mix of cattle ranching and landfills in the tropics are the main driver of the post-2006 increase, because they have expanded dramatically alongside populations in the region.

But the sharp acceleration in the past couple of years seemed to require some other source. Studies are now implicating the Sudd in South Sudan, the continent's largest swamp and a region researchers have been unable to study on the ground because of the long-term conflict in the region. Using Japan's Greenhouse Gases Observing Satellite, which measures the amount of light absorbed by methane at infrared wavelengths, Palmer and his colleagues were able to show the Sudd had grown as a methane hot spot since 2019, adding some 13 million extra tons per year to the air—more than 2% of annual global emissions. A second study, posted in late June by Harvard University researchers and submitted to *Environmental Research Letters*, finds nearly the same story, especially the surge in East Africa. When combined with smaller increases from the Amazon and the northern forests, it largely explains the observed rise in the atmosphere.

Climate change may be setting the pace of the emissions. In work published earlier this year in *Nature Communications*, Palmer and colleagues showed how East African methane emissions from 2010 to 2019, measured by satellite, synced up with a temperature pattern in the Indian Ocean that periodically warms the waters off the Horn of Africa, causing increased rainfall on land. Climate projections call for this positive phase of the Indian Ocean dipole, as it's known, to grow in strength and duration with continued global warming. If it does, Palmer says, warming will beget more methane emissions from the Sudd, which in turn could fuel more warming and rains—a positive feedback loop.

Ed Dlugokencky, an atmospheric chemist at ESRL, agrees East African wetlands may well play a big role in the methane emissions of the past 2 years. "But the question of whether it's a climate feedback yet is very difficult to answer," simply because of limited records and large yearly variations in rainfall and wetland emissions. Nisbet notes, though, that the same dynamic may be playing out across other

CREDITS: (PHOTO) MAURA AJAK/AP; (GRAPHIC) K. FRANKLIN/S

tropical wetlands. "A warming world is a wetter world in the moist tropics," Nisbet says. "We have good reason to expect, if we have a moisture and temperature increase, then biological productivity follows." Research flights over wetlands in Zambia found methane levels 10 times higher than models suggested, Nisbet and his colleagues reported in May.

The researchers who identified the East Africa link also worked to rule out another possible driver of the 2-year surge: a slowdown in the destruction of atmospheric methane. Unlike carbon dioxide, which lingers for centuries, methane only lasts a dozen years or so before it is washed out of the air, primarily by an atmospheric cleanser called the hydroxyl radical (OH). Nitrogen oxides, common pollutants from fossil fuel burning, help form OH—and nitrogen oxides declined as traffic and industry subsided during the early part of the pandemic, which should have reduced OH and allowed more methane to survive. "But we find that's not the case at all," says Daniel Jacob, an atmospheric chemist at Harvard and co-author on the second study. Matching the pandemic's estimated OH reduction in their models led to a negligible change in methane levels.

In 2021, more than 100 countries signed the Global Methane Pledge, which would cut emissions 30% from 2020 levels, primarily by plugging leaks from oil and gas infrastructure. Some scientists have even discussed removing methane from the air. But those efforts might not offset rising wetland emissions, says Benjamin Poulter, a biogeochemical modeler at NASA's Goddard Space Flight Center. "I can see a scenario where we mitigate methane—but we still see increases in atmospheric methane."

Some might seize on methane emissions as a reason for draining or developing wetlands, which are already under threat worldwide, says Eoghan Darbyshire, a researcher at the Conflict and Environment Observatory, a U.K. charity. Last year, following earlier work from Palmer that first highlighted the Sudd as a methane source, South Sudan proposed achieving its climate goals by finishing the Jonglei Canal, abandoned in the 1980s, which would divert water from the Sudd to Egypt. But draining the Sudd might just replace its methane emissions with carbon dioxide generated as newly exposed peat decomposes, while doing immeasurable damage to its ecosystem, Darbyshire says. "On the surface these seem like reasonable arguments," he says. "But if you start to think about them a little bit, they start to unravel and you're left with an overwhelming sense of uncertainty." ■

SCIENTIFIC COMMUNITY

France introduces research integrity oath

Mandatory pledge will be part of Ph.D. defense

By **Tania Rabesandratana**

Whether they're studying bioinformatics, history, or astrophysics, Ph.D. recipients in France will soon have to take an integrity oath on the day they successfully defend their thesis, in what seems to be the first national initiative of its kind. Few scientists, in France or elsewhere, believe the oath alone is likely to prevent misconduct. Nonetheless, some see it as a symbolic step in the right direction that might inspire change elsewhere.

"We had a long way to go" compared with some other countries, says Stéphanie Ruphy, director of the French Office for Research Integrity (OFIS), which helped draft the oath. France's efforts to actively promote honest, trustworthy research have sped up in recent years: introducing a national charter in 2015 laying out researchers' responsibilities, setting up OFIS in 2017, and writing procedures related to research integrity into law in 2020. Recently enacted rules, for instance, enable universities to ask for OFIS's help naming an external panel to examine alleged misconduct cases.

> **"It's a symbolic measure to affirm common values and what makes a good researcher."**
>
> **Sylvie Pommier**, France Ph.D.

The new oath is expected to become mandatory for researchers in all fields beginning their Ph.D.s or renewing their Ph.D. enrollment, starting in the fall. A draft of the oath, which had not been finalized or released as *Science* went to press, reads in part: "I pledge, to the greatest of my ability, to continue to maintain integrity in my relationship to knowledge, to my methods and to my results."

It will be mentioned in the charter signed by every Ph.D. candidate—as well as by their supervisor and institution—at the start of their doctorate, and will be taken when the Ph.D. is conferred. It won't mark entry in a specific professional body, as the Hippocratic oath does for medical doctors, nor will it be legally binding. But researchers could invoke it to bolster their opposition to dubious behavior, Ruphy says. It will also add solemnity to graduation events that, in France, often take place in nondescript rooms, without gowns or fanfare.

"It's a symbolic measure to affirm common values and what makes a good researcher," says Sylvie Pommier, president of France Ph.D., a national network of doctoral schools. Yet Pommier, who took part in the consultation about implementing the oath, and others think it should come earlier in the Ph.D. training process to instill integrity principles from the beginning of a research career.

Hugh Desmond, a philosopher of science and ethics at the University of Antwerp in Belgium, sees the oath as a good way to "strengthen a sense of professionalism among researchers, help coordinate norms, and make them public." It could "empower researchers that are lower in the hierarchy, and liberate more senior researchers," who may feel trapped by vicious career incentives and demands for quantity over quality, he adds.

Boudewijn de Bruin, an ethics professor at the University of Groningen in the Netherlands who studies oaths in professions such as accounting, is less optimistic. "I'm not against oaths in general," but their content should be detailed and specific enough to provide actual support for ethical decisions, he says. The French text, however, is brief and generic; this kind of oath will achieve "nothing," he says.

Josefin Sundin, an ecologist at the Swedish University of Agricultural Sciences who reported a case of misconduct in microplastics research (*Science*, 24 March 2017, p. 1254), says she supports the oath but is also skeptical. "The only way to improve research integrity is to promote and reward research rigor, transparency, and reproducibility over impact factor and number of publications," she says.

The oath alone won't fix these deeper problems, agrees Sundin's collaborator Dominique Roche, an ecologist and metascientist at the University of Neuchâtel in Switzerland. But it is a "positive development," he continues. "I hope other countries will follow France's lead." ■

tropical wetlands. "A warming world is a wetter world in the moist tropics," Nisbet says. "We have good reason to expect, if we have a moisture and temperature increase, then biological productivity follows." Research flights over wetlands in Zambia found methane levels 10 times higher than models suggested, Nisbet and his colleagues reported in May.

The researchers who identified the East Africa link also worked to rule out another possible driver of the 2-year surge: a slowdown in the destruction of atmospheric methane. Unlike carbon dioxide, which lingers for centuries, methane only lasts a dozen years or so before it is washed out of the air, primarily by an atmospheric cleanser called the hydroxyl radical (OH). Nitrogen oxides, common pollutants from fossil fuel burning, help form OH—and nitrogen oxides declined as traffic and industry subsided during the early part of the pandemic, which should have reduced OH and allowed more methane to survive. "But we find that's not the case at all," says Daniel Jacob, an atmospheric chemist at Harvard and co-author on the second study. Matching the pandemic's estimated OH reduction in their models led to a negligible change in methane levels.

In 2021, more than 100 countries signed the Global Methane Pledge, which would cut emissions 30% from 2020 levels, primarily by plugging leaks from oil and gas infrastructure. Some scientists have even discussed removing methane from the air. But those efforts might not offset rising wetland emissions, says Benjamin Poulter, a biogeochemical modeler at NASA's Goddard Space Flight Center. "I can see a scenario where we mitigate methane—but we still see increases in atmospheric methane."

Some might seize on methane emissions as a reason for draining or developing wetlands, which are already under threat worldwide, says Eoghan Darbyshire, a researcher at the Conflict and Environment Observatory, a U.K. charity. Last year, following earlier work from Palmer that first highlighted the Sudd as a methane source, South Sudan proposed achieving its climate goals by finishing the Jonglei Canal, abandoned in the 1980s, which would divert water from the Sudd to Egypt. But draining the Sudd might just replace its methane emissions with carbon dioxide generated as newly exposed peat decomposes, while doing immeasurable damage to its ecosystem, Darbyshire says. "On the surface these seem like reasonable arguments," he says. "But if you start to think about them a little bit, they start to unravel and you're left with an overwhelming sense of uncertainty." ■

SCIENTIFIC COMMUNITY

France introduces research integrity oath

Mandatory pledge will be part of Ph.D. defense

By **Tania Rabesandratana**

Whether they're studying bioinformatics, history, or astrophysics, Ph.D. recipients in France will soon have to take an integrity oath on the day they successfully defend their thesis, in what seems to be the first national initiative of its kind. Few scientists, in France or elsewhere, believe the oath alone is likely to prevent misconduct. Nonetheless, some see it as a symbolic step in the right direction that might inspire change elsewhere.

"We had a long way to go" compared with some other countries, says Stéphanie Ruphy, director of the French Office for Research Integrity (OFIS), which helped draft the oath. France's efforts to actively promote honest, trustworthy research have sped up in recent years: introducing a national charter in 2015 laying out researchers' responsibilities, setting up OFIS in 2017, and writing procedures related to research integrity into law in 2020. Recently enacted rules, for instance, enable universities to ask for OFIS's help naming an external panel to examine alleged misconduct cases.

> **"It's a symbolic measure to affirm common values and what makes a good researcher."**
>
> **Sylvie Pommier**, France Ph.D.

The new oath is expected to become mandatory for researchers in all fields beginning their Ph.D.s or renewing their Ph.D. enrollment, starting in the fall. A draft of the oath, which had not been finalized or released as *Science* went to press, reads in part: "I pledge, to the greatest of my ability, to continue to maintain integrity in my relationship to knowledge, to my methods and to my results."

It will be mentioned in the charter signed by every Ph.D. candidate—as well as by their supervisor and institution—at the start of their doctorate, and will be taken when the Ph.D. is conferred. It won't mark entry in a specific professional body, as the Hippocratic oath does for medical doctors, nor will it be legally binding. But researchers could invoke it to bolster their opposition to dubious behavior, Ruphy says. It will also add solemnity to graduation events that, in France, often take place in nondescript rooms, without gowns or fanfare.

"It's a symbolic measure to affirm common values and what makes a good researcher," says Sylvie Pommier, president of France Ph.D., a national network of doctoral schools. Yet Pommier, who took part in the consultation about implementing the oath, and others think it should come earlier in the Ph.D. training process to instill integrity principles from the beginning of a research career.

Hugh Desmond, a philosopher of science and ethics at the University of Antwerp in Belgium, sees the oath as a good way to "strengthen a sense of professionalism among researchers, help coordinate norms, and make them public." It could "empower researchers that are lower in the hierarchy, and liberate more senior researchers," who may feel trapped by vicious career incentives and demands for quantity over quality, he adds.

Boudewijn de Bruin, an ethics professor at the University of Groningen in the Netherlands who studies oaths in professions such as accounting, is less optimistic. "I'm not against oaths in general," but their content should be detailed and specific enough to provide actual support for ethical decisions, he says. The French text, however, is brief and generic; this kind of oath will achieve "nothing," he says.

Josefin Sundin, an ecologist at the Swedish University of Agricultural Sciences who reported a case of misconduct in microplastics research (*Science*, 24 March 2017, p. 1254), says she supports the oath but is also skeptical. "The only way to improve research integrity is to promote and reward research rigor, transparency, and reproducibility over impact factor and number of publications," she says.

The oath alone won't fix these deeper problems, agrees Sundin's collaborator Dominique Roche, an ecologist and metascientist at the University of Neuchâtel in Switzerland. But it is a "positive development," he continues. "I hope other countries will follow France's lead." ■

FEATURES

CATCHING FIRE

Past efforts to coax geothermal energy from hot, dry rock deep underground have faltered. But new techniques could crack the problem

By **Warren Cornwall**, *in Milford, Utah; Photography by* **Eric Larson**/*Flash Point SLC*

The day started inauspiciously for John McLennan, as he tried to break the curse haunting a 45-year quest to coax abundant energy from deep within Earth.

First came news of an overnight accident that left one researcher recuperating in a hotel with a sore back. Then reports trickled in that seismic sensors dangling inside holes bored deep into the Escalante Desert here were malfunctioning. Repairs were delayed by gale-force winds that whipped the sagebrush-covered hills and buffeted a drilling rig that rose 50 meters from the desert like a misplaced lighthouse. Workers were already a day behind schedule, and each day burned an additional $350,000.

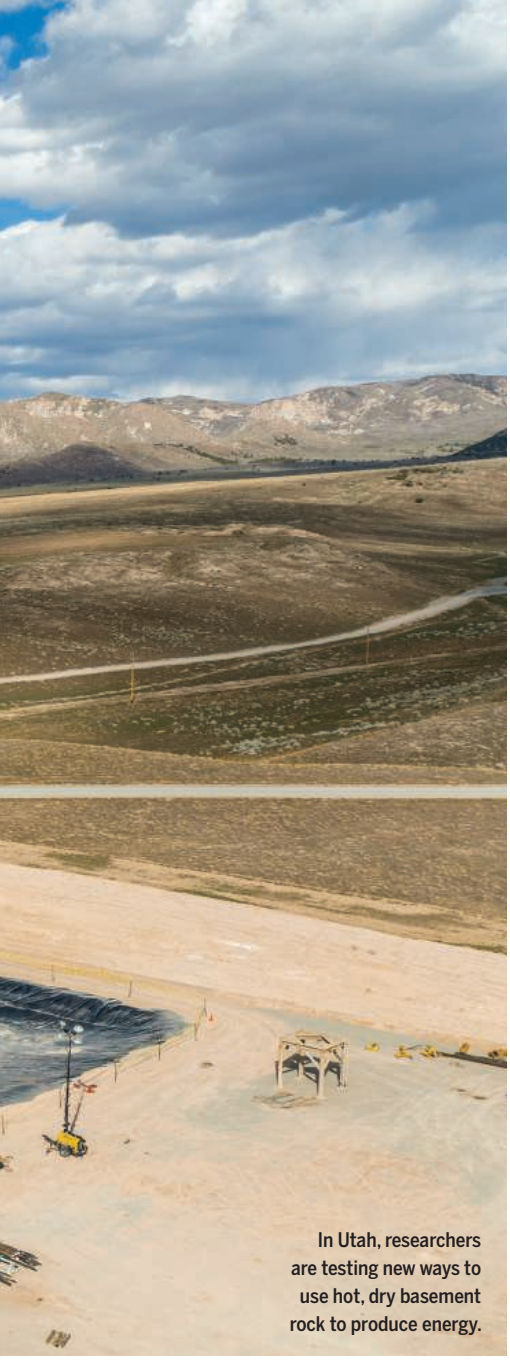
Finally, shortly before sunset, McLennan, a geomechanics engineer at the University of Utah, was ready to take a critical step in advancing a $218 million project, 4 years in the making, known as FORGE (Frontier Observatory for Research in Geothermal Energy). If successful, FORGE will help show how to transform dry, intensely hot rock found belowground all over the world into a major renewable source of electricity—and achieve a technical triumph where many others, over many years, have failed.

Gusts no longer rocked the trailer where McLennan, eyes baggy with fatigue and wearing the same brown sweater as the day before, faced five computer screens. The trailer's door opened and a co-worker—a giant of a man wearing a white hard hat—looked in. "You ready for me to go?"

"Yeah," McLennan replied. "We're ready."

With that, powerful pumps nearby sprang to life and began pushing thousands of liters of water down a hole drilled 3 kilometers into the hard granite below.

THE CONCEPT of using Earth's internal heat to generate electricity is attractively simple. Temperatures in the planet's core approach those found at the surface of the Sun, and the heat leaks outward. In places this geo-



In Utah, researchers are testing new ways to use hot, dry basement rock to produce energy.

PHOTO: ERIC LARSON/FLASH POINT SLC

thermal energy emerges at Earth's surface as molten lava, steaming vents, and hot springs. More often, however, it remains trapped in deep sediments and rock.

There's plenty of it. By one recent estimate, more than 5000 gigawatts of electricity could be extracted from heat in rock beneath the United States alone. That's nearly five times the total currently generated by all U.S. power plants. Geothermal energy is also attractive because it doesn't burn fossil fuels, isn't imported, and can run around the clock, unlike solar panels and wind turbines.

Tapping that heat, however, has proved difficult. Some nations—notably volcanically active Iceland—siphon hot groundwater to heat buildings and generate electricity. In most places, however, the rock lacks the water or the cracks needed to easily move heat to the surface. For decades engineers have sought to coax heat from this hard, dry basement rock, which can reach temperatures of more than 250°C. But those efforts have largely failed, often at huge expense—and sometimes after causing damaging earthquakes. As a result, geothermal energy provides just 0.33% of the world's electricity, little changed from 1990, according to the International Energy Agency.

Geomechanics engineer John McLennan is part of the FORGE team.

In recent years, new hope for this renewable energy source has come from an unlikely source: new technologies developed by the oil and gas industry. The same methods that have boosted fossil fuel production in the United States, such as targeted drilling and fracking—artificially fracturing deep rock with high pressure fluids—can, it's hoped, be put to work to efficiently and safely extract energy from hot, dry rock. Government agencies and private companies are pouring hundreds of millions of dollars into the approach, called enhanced geothermal systems (EGS), though it, too, has had setbacks. Now, FORGE, situated on a remote patch of land in southeastern Utah, has become a closely watched effort to demonstrate and fine-tune EGS technologies—and finally break the losing streak.

"Geothermal isn't going to work if we can't make this [EGS] work," says geologist Joseph Moore of the University of Utah, who leads FORGE. "That's really the bottom line."

FOR MCLENNAN, FORGE brings a sense of déjà vu. In 1983, when he was an engineer at an oilfield company, he worked with scientists from the Department of Energy's (DOE's) Los Alamos National Laboratory on a pioneering attempt to exploit hot, dry rock in New Mexico's Jemez Mountains. The scientists had hoped to create what amounted to artificial hot springs, by injecting water into deep fractures and then channeling the heated water back to the surface via a nearby exit well. But much of the injected water never resurfaced; researchers later concluded that they had misread the underlying geology, and the water disappeared into undetected cracks.

"It was a disappointment," McLennan recalls. And it was one of many.

Much the same thing happened decades later to a $144 million geothermal plant in Australia's arid Cooper Basin. Water pumped into the wells flowed into a previously unknown fault, and the project shut down in 2016, after just 5 years.

In some places, EGS projects had more dramatic failures, as high-pressure water injected for fracking caused existing faults to slip, setting off earthquakes. In 2006, engineers shuttered a project beneath Basel, Switzerland, after earthquakes caused minor damage. Eleven years later, a magnitude 5.5 quake struck Pohang, South Korea, killing one person, injuring dozens, and causing more than $75 million in damage. It was traced to a new, nearby EGS project where, despite a series of tremors, operators had injected fluid at high pressures near a previously unknown natural fault.

The high cost of drilling into hot, dry rock is also a challenge. Equipment designed for the softer, cooler sedimentary rock often found in oil fields falters in the extremes of hot, hard metamorphic rocks such as granite.

Today, just three EGS power plants—all near the border of France and Germany—produce electricity. In total, they generate less than 11 megawatts, enough to power about 9000 homes.

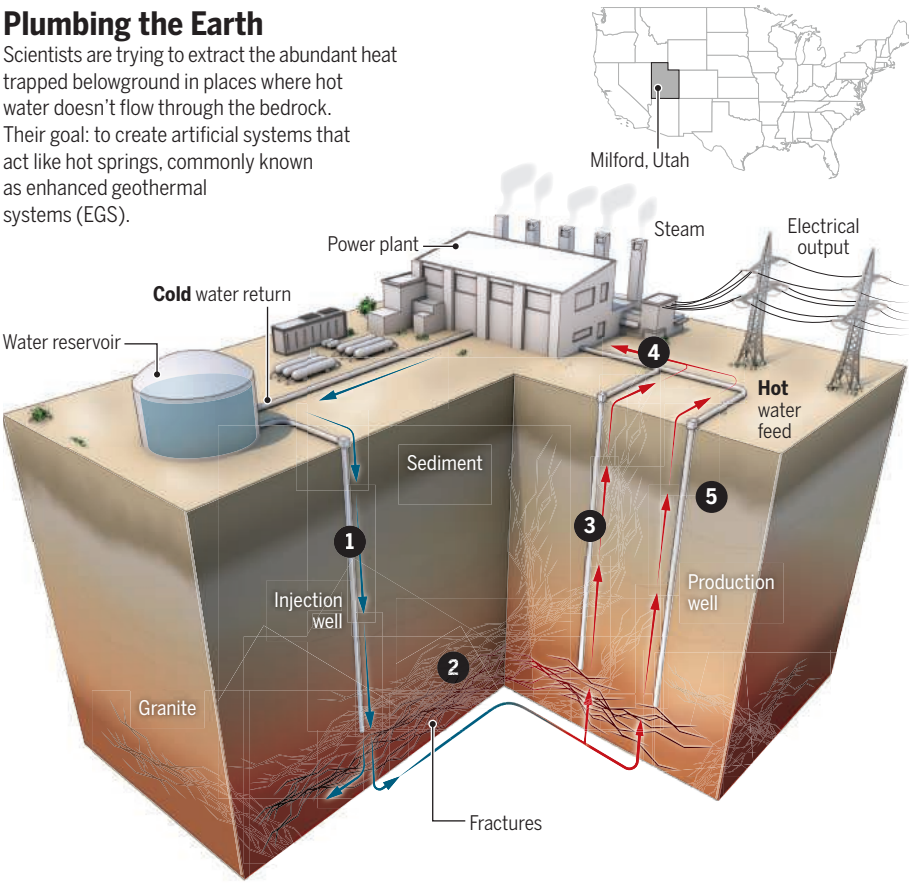
EGS "always has been fraught with technological challenges," says Jamie Beard, an attorney and executive director of the new nonprofit Project InnerSpace, which is seeking donations to help geothermal startups. And "EGS in its pure form like FORGE is hard," she adds.

Even as EGS projects have struggled, however, new techniques have emerged from the oil and gas industry. Engineers learned to drill horizontally instead of just vertically. Today they can create wells that can resemble rollercoaster routes, curving and doubling back on themselves. Sophisticated steering systems allow drillers to target their fracking to release oil and gas from veins of rock as narrow as 5 meters. The advances have prompted investors and governments to take a fresh look at EGS.

In the United States, more than two dozen geothermal companies have emerged since 2020, Beard says; that's more startups than she counts in the previous decade. In Germany, the Helmholtz Association of German Research Centers announced in June it is putting €35 million into a new underground laboratory dedicated to geothermal research in deep crystalline rock, including EGS. And DOE in April announced plans

Plumbing the Earth

Scientists are trying to extract the abundant heat trapped belowground in places where hot water doesn't flow through the bedrock. Their goal: to create artificial systems that act like hot springs, commonly known as enhanced geothermal systems (EGS).



1 In EGS, a cold fluid such as water is injected into wells deep enough to reach hot, dry rock, often kilometers down.

2 Water is heated as it flows into cracks in the rock. In EGS, high-pressure fluids are used to create new fractures or enlarge existing ones, a technique much like the fracking used in the oil and gas industry.

3 Hot water is pumped to the surface via exit wells drilled not far from the injection well. The water turns to steam as the pressure drops closer to the surface.

4 The steam is directed to turbines that generate electricity. Most of the water is reused and continues the cycle.

5 EGS-like techniques can also be used to create water that's not hot enough to efficiently generate electricity. But the artificial hot springs can be used to warm homes and commercial buildings, and drive a variety of industrial processes.

to spend $84 million on four EGS pilot projects. They'll be placed in different geological settings in the United States to study the best ways to extract heat from different types of rock.

Those plants will build on the results of FORGE, which DOE launched in 2014 with a competition to create a laboratory for honing EGS tools. In 2018, DOE announced the University of Utah and partners had won the funding to build the facility near the small railroad town of Milford, Utah, where Earth's feverish interior creeps close to the surface.

THE MILFORD VALLEY'S VENEER of vegetation is so thin that much of its geologic history is exposed like an open book. Moore has spent more than 40 years reading that tome. Earlier this year, he stood next to a low cliff of silica that runs north to south along the top of a small hill. "This is a fundamental boundary," he declared.

The wall divides what are, for the purposes of geothermal energy, two different worlds. To the east, the ground quickly rises to the flanks of the Mineral Mountains, rounded peaks speckled with granite outcroppings and juniper trees. A flat-topped mountain devoid of granite marks the top of a long-dead volcano, one of nine that testify to the heat still trapped beneath the ground in this region.

Between the cliff and the mountains, abundant hot groundwater flows close to the surface. A squat, brown building tucked into the foothills holds a 38-megawatt conventional geothermal power plant whose wells tap into that water. Just beyond, the steaming vent of a hot spring emerges next to the rock-walled ruins of a crudely built "resort" from the late 1800s. It catered to miners from the nearby Horn mine, once declared the world's richest silver deposit. At Moore's feet sat opal gemstones formed about 1600 years ago when hot springs saturated with silica spilled to the surface, leaving behind rocks candy-striped in yellow, red, and white.

West of what is known as the Opal Mound fault, the groundwater is blocked by an underground wall of solid granite that reaches temperatures of 235°C—truly hot, dry rock. FORGE's drilling rig perches above that granite, its metal skeleton dwarfing the trucks and one-story buildings clustered around it.

Starting in 2020, crews used a similar rig to drill an injection well. The completed shaft, 22 centimeters in diameter, extends for 3.3 kilometers. The well includes features that are standard in fracking operations but still cutting edge in EGS. For example, FORGE's shaft dives into the target granite at an angle that is close to horizontal—chosen to intersect with natural stresses in the rock in a way that would enable engineers to amplify tiny existing fractures.

At its deepest point, the shaft pierces rock that is 1.7 billion years old. Conventional metal drill bits struggled to cut through this stone, and the younger granite above, advancing just 3 meters per hour and frequently disintegrating. That prompted a switch to tougher drills tipped with synthetic diamonds—a first for geothermal drilling in granite. The bits sliced through the rock 10 times faster, and are "definitely a breakthrough," says Peter Meier, an engineer and CEO of the Swiss geothermal company Geo Energie Suisse, who visited FORGE earlier this year to help with seismic monitoring. "This is already a very big result of the project."

The FORGE well is lined with a steel casing that's standard in oil fracking. Such linings make it easier to use specially designed gaskets to seal off sections of pipe in which operators detonate small explosives, shattering the pipe and exposing the surrounding rock. That helps FORGE fracture rock bit by bit—another novelty in EGS.

That piecemeal approach could help EGS projects avoid fracking in seismically sensitive areas that could trigger nearby faults, says William Ellsworth, a Stanford University geophysicist who has studied drilling-induced tremors, including the Pohang quakes. But he cautions that spotting problem faults in hard basement rock is "an exceedingly difficult imaging problem."

At FORGE, moreover, the granite's high heat has crumbled the gaskets typically used in cooler oil and gas wells. It's also fried seismic sensors essential to tracking the fracking operation. So, the team has been testing special high-temperature gaskets and new monitoring tools, including fiber optic cables able to withstand the heat while detecting tiny vibrations in the earth.

GRAPHIC: C. BICKEL/*SCIENCE*

Universities, government labs, and companies are currently developing other technologies they hope to test at FORGE. One is a small device, resembling a motorized skateboard, that would drive deep into the well to open and shut "windows" in the steel casing that expose nearby rock for fracking, another technique for targeting specific regions of rock. Such "tool development … is absolutely critical" to moving EGS into the mainstream, Moore says.

ON 16 APRIL, McLennan and the FORGE team were ready for one of their first big tests: seeing whether they could pump enough water into the deepest part of the well, under enough pressure, to enlarge tiny cracks or create new ones in pockets of granite.

The equipment malfunctions and high winds had initially thrown the schedule into disarray. But by late in the day, McLennan was perched in front of his computer screens, like an air traffic controller ready to guide a plane in for a landing.

Voices crackled over radios. Beside him, Kevin England, a veteran petroleum engineer, issued short bursts of commands.

Outside, a full Moon rose over a row of water tanks, each nearly as big as a school bus. The roar of motors filled the air as powerful truck-mounted pumps moved water to the well through a spaghetti of pipes.

On McLennan's screen, numbers began to climb, tracking the water flooding into the hole, where the pressure would hit the last 60 meters of rock, left exposed without a steel shell. The engineers hoped that would allow them to create a focused, dense cloud of fractures, like an acupuncturist inserting needles into a specific nerve.

A red line crawled across the screen, marking the gradual rise in water pressure. Eventually, the line wavered, bouncing around a pressure of about 28 megapascals, more than 250 times the atmospheric pressure. The flutter was good news: It probably meant the rock was giving way. "We're getting some action," McLennan announced. "This is nice."

Over the next half hour, the signals continued to be encouraging. For the first time in days, McLennan appeared at ease. "This is beautiful," he said. "It fractures, it stops, then it propagates again."

It was hard to know exactly what was happening nearly 3 kilometers below. But Jim Rutledge, a seismologist at a nearby screen, was gathering clues. Clusters of black dots appeared on a grid, marking tiny earthquakes detected by sensors in a monitoring well half a kilometer away. The tiny tremors were no cause for alarm—just a sign that the fracking was going as planned. "We have a big cloud," Rutledge said.

At 77 minutes in, the volume of water pouring down the well had grown to 50 barrels per minute—an aspirational target some had predicted the team wouldn't reach because of the unyielding rock.

"Let's go to 60," Moore pushed.

But McLennan urged caution: "Let's stick to the plan." An hour later, the pumps went quiet.

MOORE AND MCLENNAN were buoyant the day after the April test: They had pulled off their first successful frack. Now, the FORGE team is sifting through the mountains of data collected during that test as well as two subsequent fracks at locations higher up in the same well. What they learn will shape their next steps.

At FORGE, a vast network of pipes helps move water around the drilling site.

PHOTO: ERIC LARSON/FLASH POINT SLC

A 3D map of the tiny seismic events triggered during the tests, for example, will help them decide where to drill a second well. If all goes as planned, in 2023 they will pump water down the first hole and then see what, if anything, flows back up the second.

Others are watching FORGE closely for lessons. Meier is leading plans for an EGS project in Switzerland. He hopes FORGE's technique of executing smaller, segmented fracks will point the way to reducing the risk that EGS will cause damaging earthquakes, like those that shut down his company's previous work in Basel.

Others are eager to see whether FORGE can identify ways to make EGS more commercially attractive by solving problems that today scare off would-be investors, including time-consuming drilling, broken equipment, and prolonged uncertainty over whether a well can produce hot water.

"That's where we're focusing all of our time and energy—taking away the risk from the [geothermal energy] community," says Lauren Boyd, acting director of DOE's Geothermal Technologies Office.

Others, however, see more immediate commercial promise in other strategies. Call them EGS 2.0. Beard, for example, argues for targeting softer, slightly cooler rocks at shallower depths, familiar territory for oil drilling. The approach would still rely on engineered hot springs and possibly fracking. But the formations are easier to work in, Beard says, and drillers have gained expertise from boring tens of thousands of wells into such geology.

Sage Geosystems is one company pursuing that strategy. Founded in 2020 by scientists and executives from the oil giant Shell, the Houston-based company aims to drill and frack a single well in sedimentary rock and use a set of concentric pipes to pump cool fluids into the rock and draw out hot ones. Instead of water, the firm might use liquid carbon dioxide, because it has a lower boiling point. The resulting steam would drive turbines specially designed to operate with carbon dioxide.

One longtime EGS proponent is trying a different, less technically challenging approach. Chemical engineer Jeff Tester of Cornell University helped run the Los Alamos work in the 1970s and was the lead author of a 2006 DOE report touting EGS. Now, he's overseeing a program that this summer started to drill a test well in sedimentary rock on the Cornell University campus in New York state. Although rock temperatures could top out at just 100°C, that would be enough to produce hot water to heat all the university's buildings, Tester says. And these lower temperatures are found in rock in many more places. "It doesn't have to be high temperature," he says. "That's the beautiful feature of using [cooler rock] directly for heating."

AROUND THE FORGE SITE, researchers believe the hot, dry rock just a few kilometers below holds enough heat to power a city the size of Salt Lake City. Given its mission as a testbed, however, the facility might never produce enough power to light a single bulb.

That doesn't trouble Moore. "The purpose of things like this is not to solve all the problems," Moore said the day after the first frack, as he stood on a dirt road a short distance from the drill rig. Instead, he said, FORGE's goal is to see whether it can "take [EGS] to the point where the private sector can see its viability." ■

MEDICINE

Greening the pharmacy

New measures and research are needed to limit the ecological impact of pharmaceuticals

By **Gorka Orive[1,2,3,4], Unax Lertxundi[5], Tomas Brodin[6], Peter Manning[7,8]**

The growth of pollution arising from the production and consumption of synthetic chemicals now outpaces all other environmental disruptors (e.g., rising carbon dioxide emissions) (*1*). Humans consume more pharmaceuticals than ever; in 2020, the volume of medicines used globally reached 4.5 trillion doses, and consumption continues to rise (*2*). In addition, drugs are also administered to a wide range of livestock and pets. Although pharmaceutical use brings huge benefits to human and animal health, it has also led to increased pharmaceutical pollution of ecosystems throughout the world (*3*). This pollution is likely to be exacerbated by disease epidemics and pandemics, which induce major drug spikes in aquatic ecosystems that receive wastewater, resulting in unknown ecological impacts.

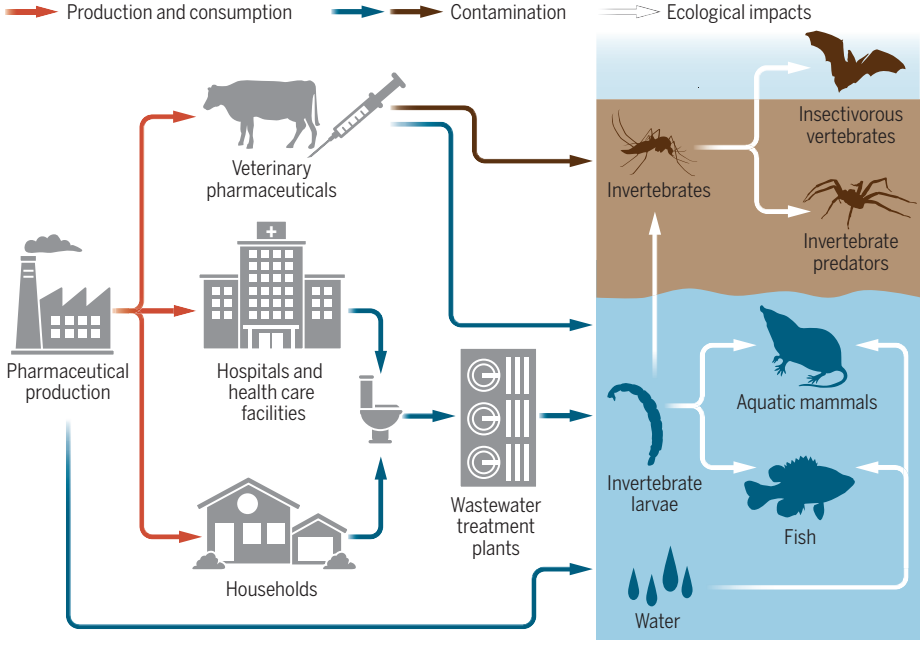
Once administered, human pharmaceuticals enter aquatic environments through wastewater; according to the United Nations Educational, Scientific and Cultural Organization (UNESCO), currently 80% of all wastewater flows into ecosystems without any form of treatment (*4*), and virtually all of the 20% that is treated still contains excreted pharmaceuticals and pathogens. As a result, pharmaceuticals are found in many different environments, on all continents, where they invoke potentially far-reaching ecological impacts for the very reason they are effective as drugs: They are molecules designed to trigger biological changes, even at extremely low concentrations. Research has shown that wild aquatic animals can accumulate pharmaceuticals to doses equivalent to those given to humans (*5*), which can result in plasma concentrations higher than the human therapeutic concentration (*6*). In addition, pharmaceutical pollutants have been shown to alter behavior, organismal fitness, and the dynamics of populations, communities, and ecosystems (see the figure) (*7*). For example, several species of vultures in Asia were driven to near extinction by exposure to the nonsteroidal anti-inflammatory drug (NSAID) diclofenac, which is used widely in livestock (*8*). Other notable examples include changes to fish behavior when exposed to antidepressants or anxiolytics, and the impact of ivermectin, an antiparasitic drug, on dung beetle diversity and ecosystem functioning.

Despite increasing evidence for potentially widespread effects, pharmaceutical pollution has not received the attention it deserves. Although there are signs that this might be changing—for example, in Europe, where the "Green Deal" commitment to tackling climate- and environment-related challenges has turned attention to pharmaceutical pollution (*9*)—there is still much to be understood. For example, investigation of the ecological impact of temporary drug spikes in surface waters has been limited, meaning that the environmental effects of increased pharmaceutical consumption during the COVID-19 pandemic is unknown.

To date, most attention on the ecological effects of pharmaceuticals has focused on endocrine-disrupting compounds (which mimic, block, or interfere with hormonal signaling) and the overuse of and attendant development of resistance to antibiotics. Although this is of huge importance for the environment and human health, the ecological impact of most other drugs remains unknown, especially within the setting of complex natural ecosystems. To date, 88% of drugs that target human proteins are lacking comprehensive environmental toxicity data (*10*). Furthermore, viruses play an important role in regulating ecosystem processes, but the potential impact of antivirals is yet to be studied in depth, despite widespread contamination of freshwater with

The life cycle of pharmaceutical drugs

Pharmaceuticals are produced by industrial facilities and are consumed in agricultural production, hospitals, health care facilities, and households. They are released into the environment, for example, through dung, urine, and wastewater, unless properly regulated and waste products are treated. Pharmaceutical contamination impacts organisms directly and indirectly through the food chain, affecting terrestrial and freshwater ecosystems. Measures to reduce the risk of drug pollution can be implemented at all stages of the life cycle.



GRAPHIC: K. HOLOSKI/*SCIENCE*

[1]Laboratory of Pharmaceutics, School of Pharmacy, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain. [2]Biomedical Research Networking Centre in Bioengineering, Biomaterials and Nanomedicine, Vitoria-Gasteiz, Spain. [3]Bioaraba, NanoBioCel Research Group, Vitoria-Gasteiz, Spain. [4]Singapore Eye Research Institute, Singapore. [5]Bioaraba Health Research Institute, Osakidetza Basque Health Service, Araba Mental Health Network, Araba Psychiatric Hospital, Pharmacy Service, Vitoria-Gasteiz, Spain. [6]Department of Wildlife, Fish, and Environmental Studies, Swedish University of Agricultural Sciences, Umeå, Sweden. [7]Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany. [8]Department of Biological Sciences, University of Bergen, Bergen, Norway. Email: gorka.orive@ehu.eus; peter.manning@uib.no

these substances. This is pressing because antiviral consumption has greatly increased owing to the COVID-19 pandemic.

Another widely prescribed group of drugs—cholesterol-lowering statins—are derived from a defense compound in molds. Although these drugs are known to affect the reproductive biology of crustaceans (*11*), they could affect all metazoans because cholesterol biosynthesis is a fundamental biological process. More broadly, many of the proteins that are targeted by human drugs are similar across the tree of life, and consequently, these drugs have the potential to affect nontarget organisms.

The potentially prolonged increases in freshwater drug concentrations that disease epidemics bring are not considered in the environmental risk assessments (ERAs) presented to health care authorities by manufacturers. Also, future water contamination is predicted to be even higher because of accelerating urbanization and the growth of megacities, which concentrate both people and pharmaceuticals. In addition, off-label use of drugs is not considered in ERAs. Currently, the standardized laboratory studies of the environmental risk of pharmaceuticals that are required by regulatory agencies focus on short-term effects of acute exposure on a small number of model organisms. This is problematic because these studies will inevitably miss many indirect effects mediated by both environmental factors (e.g., more toxic effects under conditions of heat stress) and species interactions (cascading impacts on prey, competitors, and predators) (*7*). They also only assess toxicity—a potential oversight given that most pharmaceuticals are designed to induce sublethal biological changes at low concentrations. The results of these ecotoxicology studies are then scaled up to estimate long-term and indirect effects. This likely leads to an underestimation of the true ecological impacts of many drugs; long-term (multigeneration) and ecosystem-scale experimental studies of pharmaceuticals have revealed responses that are not predictable from small-scale studies (*12*). Accordingly, the complexity and ecological relevance of pharmaceutical risk assessments need to be increased.

To truly measure the effects of drug pollution on ecosystems, studies need to include the combined impact of multiple pharmaceuticals (mixture effects) and monitor the cascading impacts of these drugs across food webs. This could be done in mesocosm studies: replicated semiclosed ecosystems that allow for a combination of ecological complexity and experimental control. Research on evolutionary effects of multigenerational exposure to pharmaceuticals is another understudied area that could be addressed in such experiments.

Ecological impacts of active pharmaceutical ingredients must also be studied in different environments throughout the world. Socio-ecological systems in tropical and arid regions are not equivalent to those in the temperate zone. However, to date, most research about the presence and impact of pharmaceuticals in the environment has been conducted in high-income, temperate-zone countries. Consequently, very little is known about the ecological effects of drug pollution in most of the world's ecosystems, especially those of the Global South, where water is often scarce and used repeatedly by people before being discharged containing high drug concentrations. For example, in 2017, ~1.7 billion treatments were delivered to >1 billion individuals in mass drug administration programs aimed at combating neglected tropical diseases. These included ivermectin and azithromycin, which have deleterious environmental effects, but the impact of this mass drug administration remains unknown (*13*). In addition, these countries also frequently experience pollution from the manufacturing of drugs that are used in high-income countries, but the ecological impacts of this are also largely unstudied. These issues must be addressed to meet the ambitious Sustainable Development Goals (SDGs) set by the United Nations, especially objective 6: clear water and sanitation (*14*).

To effectively fight drug pollution and achieve the SDG and "One Health" goal of holistically addressing health and environmental issues, a combination of source-directed and end-of-pipe measures should be implemented. The pharmaceutical industry and its customers must assess and adjust many aspects of the pharmaceutical life cycle. "Greener" drugs with lower environmental impact—for example, that are less biologically reactive or more easily eliminated from the environment—need to be designed and formulated, and drug manufacturing facilities need better wastewater management. The judicious and responsible use of drugs in both human and veterinary medicine is necessary, and nonpharmacological interventions should be prioritized when possible. Educating health care practitioners about the impact of drug pollution is also key. Redefining the concept of "rational use of drugs" by including a One-Health approach should be considered.

Drugs should be removed from wastewater before they enter the environment. The technical solutions to clean pharmaceuticals from wastewater exist (e.g., advanced oxidation with ozone), but to date, only a few countries (e.g., Switzerland) have implemented large-scale tertiary treatment of wastewater. Owing to the relatively high cost and lack of regulatory pressure, implementation of treatments that remove drugs from wastewater is still uncommon. The use of multifunctional solutions to clean multiple pollutants and pathogens from wastewater will be important when designing new or upgrading wastewater treatment plants in a cost-effective manner. Ozone-treatment of wastewater can reduce the abundance of many viruses (*15*), as well as concentrations of other pollutants that may affect aquatic biodiversity. The implementation of new techniques to treat wastewater also represents opportunities to study the chemical and biological impact of improved wastewater treatment. Such research will be invaluable for bot researchers and regulators because it woul demonstrate the sum of direct and indirec benefits of improved wastewater treatment in contrast to the theoretical benefits pre dicted from laboratory studies.

As evidence from ecotoxicology studie about the harmful effects of veterinary an human pharmaceutical pollution accumu lates, and public pressure to reduce these im pacts grows, it is important to remember th primary role of pharmacotherapy, which is t grant patients access to life-improving drugs Raising awareness of the environmental im pacts of drugs could have unforeseen societa effects, such as reluctance to take medicines guilt in doing so, and increased distrust an resentment of not only the pharmaceutica industry but pharmacotherapy more gener ally. Reduced access to effective drugs be cause of policy or price changes that reflec environmental impacts are also a possibilit However, although conflicts of interest ma be unavoidable, it is possible to limit the neg ative consequences of pharmaceuticals whil still allowing society to benefit. It is time t make green pharmacotherapy a reality. ■

REFERENCES AND NOTES

1. E. S. Bernhardt *et al.*, *Front. Ecol. Environ.* **15**, 84 (2017).
2. IMS Institute for Healthcare Informatics, "Global Medicines Use in 2020" (2015); https://rb.gy/f6qioh.
3. J. L. Wilkinson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113947119 (2022).
4. UNESCO World Water Assessment Programme, "The United Nations world water development report 2017: wastewater: the untapped resource; facts and figures" (2017); https://rb.gy/3lz9tz.
5. E. K. Richmond *et al.*, *Nat. Commun.* **9**, 4491 (2018).
6. D. Cerveny *et al.*, *Environ. Int.* **146**, 106188 (2021).
7. K. A. Kidd *et al.*, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130578 (2014).
8. J. L. Oaks *et al.*, *Nature* **427**, 630 (2004).
9. European Commission, Secretariat-General, "The European Green Deal" (2019); https://rb.gy/dzik2g.
10. L. Gunnarsson *et al.*, *Environ. Int.* **129**, 320 (2019).
11. M. M. Santos *et al.*, *Aquat. Toxicol.* **174**, 1 (2016).
12. G. Polverino *et al.*, *Proc. Biol. Sci.* **288**, 20202294 (2021).
13. G. Orive, U. Lertxundi, *Lancet* **395**, 1112 (2020).
14. S. Domingo-Echaburu, L. M. Dávalos, G. Orive, U. Lertxundi, *Sci. Total Environ.* **800**, 149412 (2021).
15. H. Wang *et al.*, *Int. J. Hyg. Environ. Health* **221**, 479 (2018).

10.1126/science.abp9554

SURFACE CHEMISTRY

A Swiss Army knife for surface chemistry

Voltage pulses offer a way to control single-molecule reactions on a surface

By **Igor Alabugin** *and* **Chaowei Hu**

To construct complex molecules and molecular devices, tiny, atomic-sized objects must be brought together and connected in a precise way. For better or for worse, this daunting task is still mostly done in a manner likened to putting Lego blocks in a washing machine and hoping that the quintillions of molecules somehow end up assembling themselves into the desired product, either by complete chance or under the guidance of other molecular-sized objects—i.e., catalysts. On page 298 of this issue, Albrecht *et al.* (*1*) show how a single molecule can be transformed into three distinct products depending on the voltage pulses from the tip of a scanning tunneling microscope (STM). Notably, the three products can be repeatedly interconverted with a high degree of control.

On-demand interconversion of molecules, or switching, can be done in many ways, such as chemically, photochemically, or electrochemically, depending on the chemical system and the intended goal. Recently, controlled transformations of individual molecules on surfaces have become possible (*2*–*6*). Although switching often relies on well-known processes, Albrecht *et al.* describe a distinct network of reactions that connects three exotic species, each of which would have limited stability under ambient conditions. These species defy the usual chemistry logic but are sufficiently stable on a sodium chloride (NaCl) surface to take part in multiple transformations and measurements, as demonstrated in 440 different reactions performed on five individual molecules. These processes revealed unusual chemistry that offers a large degree of control of reactivity and bond formation.

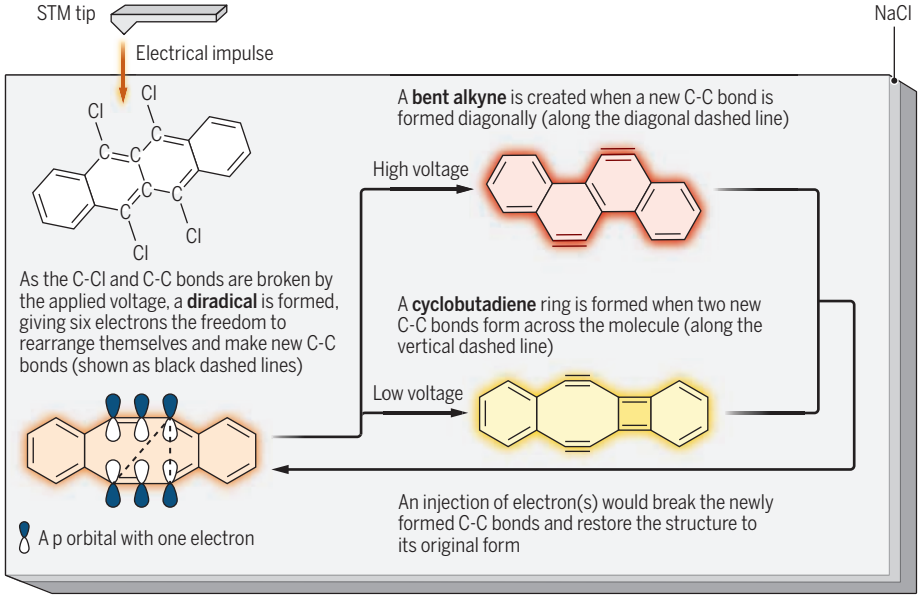
The sequence of transformations was set up by removing four chlorine atoms from the tetracyclic core to create the starting material, using an atomically sharp STM tip positioned above the particular molecule. As the chlorine atoms are removed, radical centers (i.e., an atom with one or more unpaired electrons) are created on the carbon atoms of the four carbon-chlorine (C-Cl) bonds. Two of the radical centers will then recouple by breaking one of the C-C bonds in the ring, but, after all the Cl atoms have departed, the system is still left with two unpaired electrons (i.e., a diradical). In this state, the two radical centers that are spread over two three-carbon sets of atoms are placed precariously, facing each other.

The usual chemical logic suggests that such a diradical would be transient and that the two unpaired electrons would immediately reconnect to make a C-C bond. However, in the unusual system reported by Albrecht *et al.*, the situation is quite different because of the presence of chemical frustration—i.e., the inability for the system to satisfy bonding requirements because of structural constraints (see the figure). Even though there are several ways to couple the electrons, the system is constrained structurally from forming the bond. The available choices are poor because one leads to an antiaromatic cyclobutadiene ring and the other forms a highly bent alkyne with a distorted C-C triple bond. Although the unpaired electrons can succeed in forming the C-C bonds needed for the products, the bonds are relatively weak and can be broken and remade at will by an applied voltage. Out of the two viable options, the unstable four-membered cyclobutadiene ring is preferred by the system because it is easier for the structure to form compared with the six-membered rings in the bent alkyne, which requires a greater distortion from its original geometry. However, the alkyne becomes the main product when a higher voltage is applied. This interplay between how quickly the product can be formed and its stability provides a way to selectively make either cyclobutadiene or bent alkyne by controlling the input voltage.

The reversion of two products back to the diradical requires energy and, hence, is unfavorable. Albrecht *et al.* made the back switching possible by inverting the relative stability of the diradical versus the cyclobutadiene and the alkyne by injecting electrons into the diradical using a voltage pulse. Because diradicals have a high likelihood of gaining an electron (*7*, *8*), their negatively charged states are suddenly more stable relative to the anionic states of the other two products. This change in the molecular energy landscape facilitates the conversion of anionic

Molecular shape-shifting enabled by electrochemistry

Albrecht *et al.* show how multiple reactions can be initiated and controlled by applying a voltage using a scanning tunneling microscope (STM) and demonstrate a way to interconvert between three distinct molecules.



GRAPHIC: A. MASTIN/*SCIENCE*

Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA.
Email: alabugin@chem.fsu.edu

forms of cyclobutadiene and bent alkyne back to radical anions. The experiments show that at least two electrons are needed for this to happen—one to provide energy for the reaction and one that is retained to open the ring. Furthermore, an observation of a transient dianion suggests that the extra electron may serve as a catalyst for the reaction (*9*, *10*) because the removal of the extra electron restores the neutrality and returns the system to its original state.

The control of voltage in conjunction with the possibility of electron injection enables a fully controlled interconversion between three isomeric, electronically unusual, and inherently unstable molecules. Albrecht *et al.* put a spotlight on these exotic species and illustrated how C-C bonds in carbon-rich species with weakly coupled electrons can be formed, broken, and reformed at will. Lessons from these newly discovered surface chemistry processes may be relevant for electron-catalyzed chemical reactions in chemistry and biology and may illuminate viable shortcuts for more efficient synthesis.

This molecular system can on-demand become one of three exotic chemical creatures, transforming between a diradical, an antiaromatic ring, and a highly bent alkyne. The potential to interact with a different set of partners makes this shape-shifting molecular system a Swiss Army knife with three distinct and useful chemical tools. For example, the diradical may participate in redox and radical reactions, the cyclobutadiene system can provide a potential binding site for a cationic transition metal, and the strained alkynes units can be used for various reactions such as cycloaddition (*11*) or nucleophilic attack (*12*, *13*). Future opportunities may arise from using the differences in electronic structures of the three species in molecular electronics, such as in molecular logic gates (*14*, *15*). ■

REFERENCES AND NOTES

1. F. Albrecht *et al.*, *Science* **377**, 298 (2022).
2. D. G. de Oteyza *et al.*, *Science* **340**, 1434 (2013).
3. N. Pavliček *et al.*, *Nat. Chem.* **7**, 623 (2015).
4. A. Riss *et al.*, *Nat. Chem.* **8**, 678 (2016).
5. B. Schuler *et al.*, *Nat. Chem.* **8**, 220 (2016).
6. N. Pavliček *et al.*, *Nat. Nanotechnol.* **12**, 308 (2017).
7. P. W. Peterson *et al.*, *J. Am. Chem. Soc.* **138**, 15617 (2016).
8. L. A. Hammad, P. G. Wenthold, *J. Am. Chem. Soc.* **125**, 10796 (2003).
9. A. Studer, D. P. Curran, *Nat. Chem.* **6**, 765 (2014).
10. M. A. Syroeshkin *et al.*, *Angew. Chem. Int. Ed.* **58**, 5532 (2019).
11. N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **126**, 15046 (2004).
12. X. Xiao, T. R. Hoye, *J. Am. Chem. Soc.* **141**, 9813 (2019).
13. T. Harris, I. V. Alabugin, *Mendeleev Commun.* **29**, 237 (2019).
14. A. Credi, V. Balzani, S. J. Langford, J. F. Stoddart, *J. Am. Chem. Soc.* **119**, 2679 (1997).
15. S. Erbas-Cakmak, D. A. Leigh, C. T. McTernan, A. L. Nussbaumer, *Chem. Rev.* **115**, 10081 (2015).

10.1126/science.abq2622

NEUROSCIENCE

A cellular switchboard in memory circuits

Neurogliaform cells can direct the flow of information through the hippocampus

By **Michael T. Craig**[1] *and* **Jonathan Witton**[2]

The hippocampus is a brain region that is associated with memory. However, the hippocampus does not function alone, but rather operates within a wider network of brain regions (the extended memory network) including, among other areas, the prefrontal and entorhinal cortices and midline thalamic nuclei, such as nucleus reuniens (NRe) (*1*). Communication between these brain regions is important for many aspects of memory acquisition and consolidation, as well as for spatial navigation and decision-making. There are multiple routes through which information can flow through the extended memory network, with direct and indirect pathways converging on the hippocampus. The mechanisms by which information flow through these different pathways is prioritized have remained largely unknown. On page 324 of this issue, Sakalar *et al.* (*2*) report a cellular mechanism of information routing through the hippocampus.

Communication between neurons is thought to be enabled by neuronal oscillations—waves of rhythmic electrical activity that facilitate neural dialogue by creating temporal windows in which neuronal firing can be synchronized (*3*). By convention, neuronal oscillations are grouped into different frequency bands, with each band associated with specific cognitive processes. For example, theta oscillations occur at ~5 to 12 Hz and are associated with spatial navigation, whereas gamma oscillations occur between ~30 and 140 Hz and are associated with memory or high cognitive load (*4*). Gamma oscillations, often occurring alongside theta oscillations, can be further parsed into distinct subbands driven by different cellular mechanisms (*4*). This can be observed in the CA1 region of the hippocampus, where different types of gamma oscillation are found—specifically, a slow gamma oscillation ($gamma_S$; ~40 Hz) driven by input from neighboring CA3 and a faster midfrequency gamma oscillation ($gamma_M$; ~75 Hz) driven by input from the entorhinal cortex (*5*). $Gamma_M$ may be involved in the encoding of memory, whereas $gamma_S$ is likely to be important for memory retrieval (*5*).

Pyramidal neurons form the main computational unit of the hippocampus, with those in CA1 integrating inputs from multiple sources and sending signals to the subiculum and beyond through the generation of action potentials (also called spiking). Inputs to CA1 from CA3 arrive in stratum radiatum of the hippocampus, whereas those from the entorhinal cortex terminate in stratum lacunosum-moleculare of the hippocampus, providing an anatomical segregation of these different information streams (see the figure). There is also a functional segregation of CA3- and entorhinal-driven gamma oscillations: These different types of oscillation occur at different phases of the CA1 theta oscillation, potentially presenting a circuit-level mechanism that prevents the processes driven by different information streams (e.g., memory encoding versus retrieval) from interfering with each other (*5*). Whether the inputs to an individual pyramidal cell in CA1 can be actively switched between these different information streams has been unknown.

Within the hippocampus, inhibitory interneurons make up a diverse family of neurons, using γ-aminobutyric acid (GABA) as their neurotransmitter, with multiple subtypes providing exquisite temporal control over the spiking of excitatory pyramidal cells and other inhibitory interneurons (*6*). Neuronal oscillations are typically generated through a precisely coordinated balance between excitation and inhibition (*6*). Neurogliaform cells are an abundant class of inhibitory interneurons that reside in and project dense axonal arbors throughout stratum lacunosum-moleculare of the hippocampus. They are therefore well placed to inhibit the distal apical dendrites of CA1 pyramidal cells (*7*), but understanding their role in hippocampal information processing has remained elusive.

Sakalar *et al.* provide evidence that neurogliaform cells play an important role in

[1]School of Psychology and Neuroscience, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. [2]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK. Email: mick.craig@glasgow.ac.uk

forms of cyclobutadiene and bent alkyne back to radical anions. The experiments show that at least two electrons are needed for this to happen—one to provide energy for the reaction and one that is retained to open the ring. Furthermore, an observation of a transient dianion suggests that the extra electron may serve as a catalyst for the reaction (*9*, *10*) because the removal of the extra electron restores the neutrality and returns the system to its original state.

The control of voltage in conjunction with the possibility of electron injection enables a fully controlled interconversion between three isomeric, electronically unusual, and inherently unstable molecules. Albrecht *et al.* put a spotlight on these exotic species and illustrated how C-C bonds in carbon-rich species with weakly coupled electrons can be formed, broken, and reformed at will. Lessons from these newly discovered surface chemistry processes may be relevant for electron-catalyzed chemical reactions in chemistry and biology and may illuminate viable shortcuts for more efficient synthesis.

This molecular system can on-demand become one of three exotic chemical creatures, transforming between a diradical, an antiaromatic ring, and a highly bent alkyne. The potential to interact with a different set of partners makes this shape-shifting molecular system a Swiss Army knife with three distinct and useful chemical tools. For example, the diradical may participate in redox and radical reactions, the cyclobutadiene system can provide a potential binding site for a cationic transition metal, and the strained alkynes units can be used for various reactions such as cycloaddition (*11*) or nucleophilic attack (*12*, *13*). Future opportunities may arise from using the differences in electronic structures of the three species in molecular electronics, such as in molecular logic gates (*14*, *15*). ■

REFERENCES AND NOTES

1. F. Albrecht *et al.*, *Science* **377**, 298 (2022).
2. D. G. de Oteyza *et al.*, *Science* **340**, 1434 (2013).
3. N. Pavliček *et al.*, *Nat. Chem.* **7**, 623 (2015).
4. A. Riss *et al.*, *Nat. Chem.* **8**, 678 (2016).
5. B. Schuler *et al.*, *Nat. Chem.* **8**, 220 (2016).
6. N. Pavliček *et al.*, *Nat. Nanotechnol.* **12**, 308 (2017).
7. P. W. Peterson *et al.*, *J. Am. Chem. Soc.* **138**, 15617 (2016).
8. L. A. Hammad, P. G. Wenthold, *J. Am. Chem. Soc.* **125**, 10796 (2003).
9. A. Studer, D. P. Curran, *Nat. Chem.* **6**, 765 (2014).
10. M. A. Syroeshkin *et al.*, *Angew. Chem. Int. Ed.* **58**, 5532 (2019).
11. N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **126**, 15046 (2004).
12. X. Xiao, T. R. Hoye, *J. Am. Chem. Soc.* **141**, 9813 (2019).
13. T. Harris, I. V. Alabugin, *Mendeleev Commun.* **29**, 237 (2019).
14. A. Credi, V. Balzani, S. J. Langford, J. F. Stoddart, *J. Am. Chem. Soc.* **119**, 2679 (1997).
15. S. Erbas-Cakmak, D. A. Leigh, C. T. McTernan, A. L. Nussbaumer, *Chem. Rev.* **115**, 10081 (2015).

10.1126/science.abq2622

NEUROSCIENCE

A cellular switchboard in memory circuits

Neurogliaform cells can direct the flow of information through the hippocampus

By **Michael T. Craig**[1] *and* **Jonathan Witton**[2]

The hippocampus is a brain region that is associated with memory. However, the hippocampus does not function alone, but rather operates within a wider network of brain regions (the extended memory network) including, among other areas, the prefrontal and entorhinal cortices and midline thalamic nuclei, such as nucleus reuniens (NRe) (*1*). Communication between these brain regions is important for many aspects of memory acquisition and consolidation, as well as for spatial navigation and decision-making. There are multiple routes through which information can flow through the extended memory network, with direct and indirect pathways converging on the hippocampus. The mechanisms by which information flow through these different pathways is prioritized have remained largely unknown. On page 324 of this issue, Sakalar *et al.* (*2*) report a cellular mechanism of information routing through the hippocampus.

Communication between neurons is thought to be enabled by neuronal oscillations—waves of rhythmic electrical activity that facilitate neural dialogue by creating temporal windows in which neuronal firing can be synchronized (*3*). By convention, neuronal oscillations are grouped into different frequency bands, with each band associated with specific cognitive processes. For example, theta oscillations occur at ~5 to 12 Hz and are associated with spatial navigation, whereas gamma oscillations occur between ~30 and 140 Hz and are associated with memory or high cognitive load (*4*). Gamma oscillations, often occurring alongside theta oscillations, can be further parsed into distinct subbands driven by different cellular mechanisms (*4*). This can be observed in the CA1 region of the hippocampus, where different types of gamma oscillation are found—specifically, a slow gamma oscillation ($gamma_S$; ~40 Hz) driven by input from neighboring CA3 and a faster midfrequency gamma oscillation ($gamma_M$; ~75 Hz) driven by input from the entorhinal cortex (*5*). $Gamma_M$ may be involved in the encoding of memory, whereas $gamma_S$ is likely to be important for memory retrieval (*5*).

Pyramidal neurons form the main computational unit of the hippocampus, with those in CA1 integrating inputs from multiple sources and sending signals to the subiculum and beyond through the generation of action potentials (also called spiking). Inputs to CA1 from CA3 arrive in stratum radiatum of the hippocampus, whereas those from the entorhinal cortex terminate in stratum lacunosum-moleculare of the hippocampus, providing an anatomical segregation of these different information streams (see the figure). There is also a functional segregation of CA3- and entorhinal-driven gamma oscillations: These different types of oscillation occur at different phases of the CA1 theta oscillation, potentially presenting a circuit-level mechanism that prevents the processes driven by different information streams (e.g., memory encoding versus retrieval) from interfering with each other (*5*). Whether the inputs to an individual pyramidal cell in CA1 can be actively switched between these different information streams has been unknown.

Within the hippocampus, inhibitory interneurons make up a diverse family of neurons, using γ-aminobutyric acid (GABA) as their neurotransmitter, with multiple subtypes providing exquisite temporal control over the spiking of excitatory pyramidal cells and other inhibitory interneurons (*6*). Neuronal oscillations are typically generated through a precisely coordinated balance between excitation and inhibition (*6*). Neurogliaform cells are an abundant class of inhibitory interneurons that reside in and project dense axonal arbors throughout stratum lacunosum-moleculare of the hippocampus. They are therefore well placed to inhibit the distal apical dendrites of CA1 pyramidal cells (*7*), but understanding their role in hippocampal information processing has remained elusive.

Sakalar *et al.* provide evidence that neurogliaform cells play an important role in

[1]School of Psychology and Neuroscience, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. [2]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK. Email: mick.craig@glasgow.ac.uk

routing information flow to CA1 pyramidal cells. Using mice running in a virtual reality environment, the authors recorded spiking and neuronal oscillations across CA1 alongside juxtacellular recordings of putative and anatomically verified neurogliaform cells. They report that the timing of neurogliaform cell spiking is strongly modulated by $gamma_M$ and that neurogliaform cell firing can uncouple CA1 pyramidal cells from $gamma_M$ oscillations, indicating temporary disconnection of pyramidal cells from entorhinal input. GABAergic inhibition is usually considered to decrease cellular excitability, but neurogliaform cells appear to suppress $gamma_M$ modulation of CA1 pyramidal cell activity without changing the overall firing rate of the pyramidal cells, suggesting that neurogliaform cell–mediated inhibition is restricted to the most distal compartments of the pyramidal cell's apical dendrites.

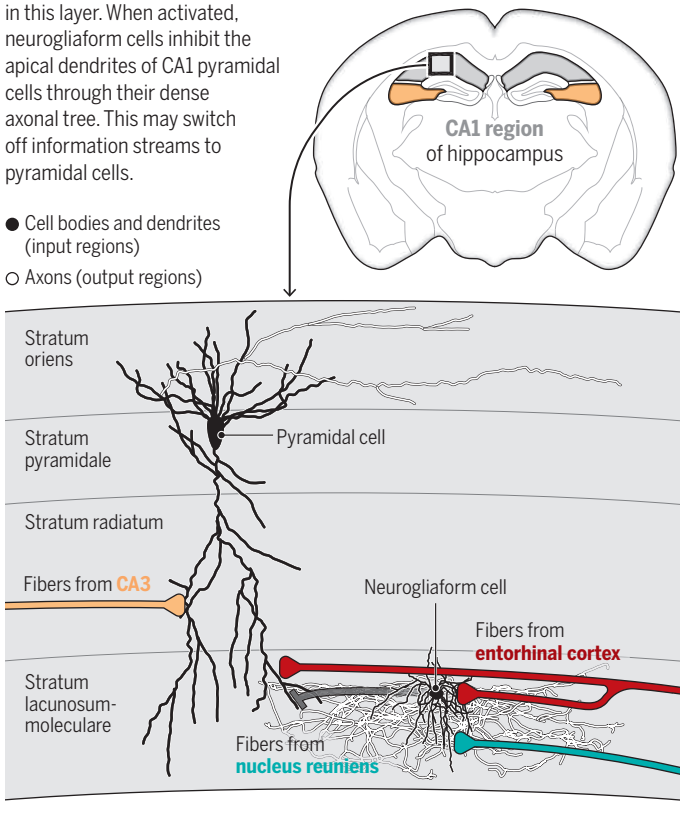
How does neurogliaform-driven uncoupling of CA1 pyramidal cells from entorhinal cortex inputs relate to information processing in the hippocampus? Numerous aspects of cognition, such as memory and decision-making, require information to be communicated between the hippocampus, prefrontal cortex, and entorhinal cortex. This communication is thought to be facilitated by the synchronization of neuronal oscillations across brain regions. For example, temporal coupling between CA1 and prefrontal theta and gamma oscillations (*8*, *9*) and between CA1 and entorhinal gamma oscillations (*10*) occurs during decision-making in spatial working memory tasks; these studies were carried out in rodents, but the connectivity and function of these circuits are conserved across mammalian species. Although synchrony between the prefrontal cortex and CA1 is important for memory and decision-making, no direct excitatory connection exists between these areas, although an inhibitory projection from the prefrontal cortex to interneurons in the hippocampus was recently described (*11*). Rather, excitatory glutamatergic prefrontal cortex input to CA1 is relayed through the thalamic NRe, which also projects to the entorhinal cortex. Axons from NRe terminate in CA1 in stratum lacunosum-moleculare alongside those from the entorhinal cortex (*12*). Neurogliaform cells receive inputs from both NRe and the entorhinal cortex (*13*), and preliminary data suggest that NRe preferentially targets neurogliaform cells in CA1 while largely or entirely avoiding pyramidal cells (*14*).

Glutamatergic neurotransmission excites neurons through the activation of two main receptor subtypes, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) and *N*-methyl-D-aspartate (NMDA) receptors. AMPA receptors are activated during low and high levels of activity, but NMDA receptors only become active after a large excitatory stimulus. Although AMPA receptor–mediated synaptic currents from NRe onto neurogliaform cells are too small to drive spiking, NRe-to–neurogliaform cell synapses have a much larger NMDA receptor–mediated component (*13*, *14*). This implies that NRe provides a stronger drive to neurogliaform cells when they are simultaneously activated by another input, such as that arising from the entorhinal cortex.

Could the function of NRe input to CA1 neurogliaform cells be to assist the prefrontal cortex in selectively disconnecting pyramidal cells from entorhinal input, thereby filtering entorhinal signaling such that specific CA1 pyramidal cells are recruited? This is tentatively supported by evidence that the prefrontal cortex provides an important contextual signal to CA1 that allows the learning of new rules when the hippocampus is receiving familiar spatial information (*15*). An analogous experience for humans would be relearning the location of your favorite foods after your local grocery store has rearranged product locations. Alternatively, could NRe and the entorhinal cortex act synergistically through neurogliaform cells to shut down entorhinal input to CA1, thereby creating a permissive state for CA1 pyramidal cells to be driven by the trisynaptic entorhinal–dentate gyrus–CA3 pathway that arrives via the Schaffer collaterals in stratum radiatum? This could be an important circuit for switching the hippocampus from a state of encoding memory to a state where retrieval can occur (switching between reading and writing modes of operation). Or perhaps neurogliaform cells support top-down prefrontal cortex inhibition of ongoing entorhinal-CA1 communication in preparation for new information to be transmitted to CA1 via entorhinal-to–dentate gyrus or entorhinal-to-CA1 projections. Such a hypothetical function could provide a cellular substrate for the cognitive flexibility required to allow rapid changes of behavior while executing a task, such as discovering a hidden predator when foraging for food. Determining the functional role for this neurogliaform cell–mediated suppression of entorhinal modulation of CA1 pyramidal cell activity will present exciting challenges for those investigating hippocampus-dependent cognition. ■

Information flow through pyramidal cells

CA1 pyramidal cells can be activated by fibers from CA3 through its dendrites in stratum radiatum, whereas inputs from the entorhinal cortex or nucleus reuniens arrive in stratum lacunosum-moleculare and can target the dendrites of either pyramidal cells or neurogliaform cells in this layer. When activated, neurogliaform cells inhibit the apical dendrites of CA1 pyramidal cells through their dense axonal tree. This may switch off information streams to pyramidal cells.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. J. P. Aggleton, *Proc. R. Soc. B.* **281**, 20140565 (2014).
2. E. Sakalar *et al.*, *Science* **377**, 324 (2022).
3. P. Fries, *Neuron* **88**, 220 (2015).
4. G. Buzsáki, X.-J. Wang, *Annu. Rev. Neurosci.* **35**, 203 (2012).
5. L. L. Colgin *et al.*, *Nature* **462**, 353 (2009).
6. K. A. Pelkey *et al.*, *Physiol. Rev.* **97**, 1619 (2017).
7. L. Overstreet-Wadiche, C. J. McBain, *Nat. Rev. Neurosci.* **16**, 458 (2015).
8. H. L. Hallock *et al.*, *J. Neurosci.* **36**, 8372 (2016).
9. M. W. Jones, M. A. Wilson, *PLOS Biol.* **3**, e402 (2005).
10. J. Yamamoto *et al.*, *Cell* **157**, 845 (2014).
11. R. Malik *et al.*, *Cell* **185**, 1602 (2022).
12. F. G. Wouterlood *et al.*, *J. Comp. Neurol.* **296**, 179 (1990).
13. R. Chittajallu *et al.*, *Nat. Commun.* **8**, 152 (2017).
14. L. Andrianova *et al.*, bioRxiv 10.1101/2021.09.30.462517 (2021).
15. K. G. Guise, M. L. Shapiro, *Neuron* **94**, 183 (2017).

10.1126/science.add2681

ICE CHEMISTRY

Seeing how ice breaks the rule

Basic defects in ice monolayers are seen using a microscope

By **Yoshiaki Sugimoto**

The process of liquid adsorption onto solid surfaces can be observed everywhere, from the sweat stain on your shirt to the molecular clouds in outer space. This process, known as wetting, plays an important role in various chemical reactions and is at the core of many processes, such as corrosion and catalysis. For example, protons in the water adsorption layer are responsible for electrical conduction in electrochemical reactions on battery electrodes. Obtaining information on the location and dynamics of the protons is an important but challenging task. On page 315 of this issue, Tian *et al*. (*1*) accomplish this by preparing monolayers of water molecules on metal surfaces and using atomic force microscopy (AFM) to visualize the individual excess protons introduced onto the monolayers. They identify two qualitatively different cations with excess protons, and the abundance ratio of the two cations vary with the metal substrates.

The wetting of solid surfaces has long been investigated because of its role in a wide range of processes, such as in friction, catalysis, and electrochemical reactions. In recent years, hydrogen energy has become a promising means to decarbonize energy systems. The hydrogen fuel cell, in which the electrolysis of water molecules generates hydrogen, is a key technology for achieving this goal. Understanding the adsorption of water molecules on the electrode surface and the behavior of the protons in the water adsorption layer is needed to improve this reaction. Although many studies have been conducted, including spectroscopic experiments and simulations, a microscopic understanding of the protons in the water adsorption layer has been lacking.

Tian *et al*. used a well-defined monolayer of water molecules formed on a metal surface as a model system for water-wetted solid surfaces. Similar to crystalline ice, water molecules in a monolayer on a metal surface obey the "ice rules," also known as the Bernal-Fowler rules (*2*), and form a network structure of hydrogen bonds. During this process, one water molecule directs its hydrogen atoms toward the oxygen atoms of neighboring water molecules and in turn receives hydrogen atoms from other neighboring water molecules. In other words, the structure of the water molecule itself is maintained, and adjacent water molecules are linked together by hydrogen bonds to form a network. In an impurity-free hydrogen bonding network, electrical conduction occurs through the transfer of protons while maintaining the crystal lattice with the oxygen atoms of the water molecules. Thus, misalignments—orientation defects—and excess protons of water molecules that locally break the ice rules are important to consider (see the figure).

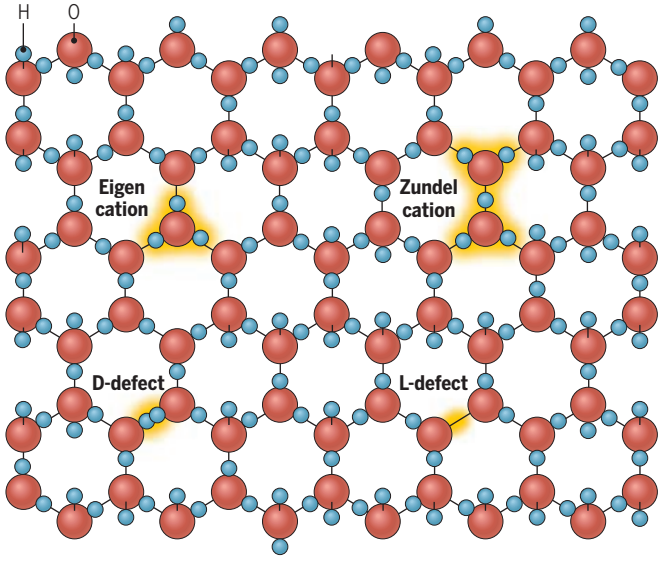
Orientation defects include D-defects (from "doppel" in German, meaning "double"), in which two hydrogen atoms of water molecules face each other, and L-defects (from "leer" in German, meaning "empty"), in which their hydrogen atoms face away from each other. The cations with excess protons are also important for generating molecular hydrogen. This includes an Eigen cation, in which H_3O^+ makes three hydrogen bonds, and a Zundel cation, in which an excess proton is equally shared between two water molecules.

Microscopic investigation of the behavior of excess protons requires the use of microscopy. AFM can be used to perform high-resolution local structure analysis. It is a microscope that scans a surface with a tip (*3*) similar to how a fingertip can sense scratches on a tabletop. By operating the AFM in an ultrahigh vacuum at a low temperature of 5 K, the tip apex can be decorated with a single molecule, such as a carbon monoxide, for ultrahigh-resolution imaging (*4*). Single-molecule chemistry based on submolecular resolution observation of organic molecules adsorbed on surfaces has garnered interest (*5*), and the technique allows the nondestructive observation of weak hydrogen bonding networks. Before the study of Tian *et al*., it has been reported that hydrogen bonding networks on metal surfaces can be visualized with high resolution by using AFM (*6*–*9*), enough to reveal local structures such as D-defects (*6*) and crystal growth front (*9*).

Tian *et al*. prepared Eigen and Zundel cations by introducing protons into the hydrogen bonding networks of water monolayers. By using AFM, they could distinguish between water molecules and hydrated protons. By changing the number of introduced hydrogen atoms, the authors created Eigen and Zundel cations separately. In both cases, calculations showed that water molecules in hydrogen bonding with the excess protons were slightly elevated from the monolayers, as corroborated by the AFM image. Furthermore, the two types of cations could be reversibly switched by applying a voltage pulse between the tip and the metal surface beneath the ice layer. Such a conversion requires the protons to move between the network of water molecules and the metal

Visualizing basic defects in ice

Following the ice rule, a monolayer of water molecules would link by their hydrogen bonds to form a structure upon freezing. There are four typical defects that break this rule: the Eigen cation, the Zundel cation, the D-defect, and the L-defect. Tian *et al*. used atomic force microscopy and visualized the former two.



An **Eigen cation** sees a water molecule with an excess proton, creating a H_3O^+ ion.

A **Zundel cation** has an excess proton shared between two water molecules (creating two $H_{25}O^{0.5+}$ ions).

A **D-defect** has two hydrogen atoms from neighboring water molecules sharing a single bond.

An **L-defect** has the hydrogen atoms from neighboring molecules facing away from each other.

GRAPHIC: N. DESAI/*SCIENCE*

Department of Advanced Materials Science, University of Tokyo, Tokyo, Japan. Email: ysugimoto@k.u-tokyo.ac.jp

surface. This information is valuable for understanding the elementary processes involved in the hydrogen evolution reaction, by explaining the path taken by the protons going from the water molecule network to the substrate. Furthermore, Tian *et al.* clarified that the abundance ratio of Eigen and Zundel cations differs between platinum and gold substrates. Because platinum is a commonly used electrode material, Tian *et al.* highlight the importance of understanding the substrate dependence of the hydrogen evolution mechanism, including how proton transfer to the substrate may be affected by factors such as the lattice constant and surface potential of the substrate.

Tian *et al.* provide a direct observation of excess protons in the water molecule network. In addition to information on atomic positions obtained with AFM, the quantitative measurements of local charge distributions will also be useful for further analyzing more complex systems, such as hydrogen bonding networks that contain injected heterogeneous ions. Measuring the electrostatic force with AFM allows visualization of local charges within a single organic molecule (*10*). This approach will help determine the directionality of hydrogen bonds and identify ionic species.

High-resolution experiments have been conducted in ultrahigh vacuum at low temperature, whereas hydration structures have been observed at the solid-liquid interface with AFM at room temperature (*11*). The hydration structure is visualized as a time-averaged image of dynamically moving water molecules. It is important to conduct research that bridges the gap between static but high-resolution observations in ultrahigh vacuum and time-averaged observations in liquid that are close to more realistic conditions. In the future, real-time observation of proton transfers in the water adsorption layers in relation to electrochemical reactions will be a key technology relevant to various fields. ■

REFERENCES AND NOTES

1. Y. Tian *et al.*, *Science* **377**, 315 (2022).
2. T. Bartels-Rausch *et al.*, *Rev. Mod. Phys.* **84**, 885 (2012).
3. G. Binnig, C. F. Quate, C. Gerber, *Phys. Rev. Lett.* **56**, 930 (1986).
4. L. Gross, F. Mohn, N. Moll, P. Liljeroth, G. Meyer, *Science* **325**, 1110 (2009).
5. L. Gross *et al.*, *Angew. Chem. Int. Ed.* **57**, 3888 (2018).
6. A. Shiotari, Y. Sugimoto, *Nat. Commun.* **8**, 14313 (2017).
7. J. Peng *et al.*, *Nat. Commun.* **9**, 122 (2018).
8. A. Shiotari, Y. Sugimoto, H. Kamio, *Phys. Rev. Mater.* **3**, 093001(R) (2019).
9. R. Ma *et al.*, *Nature* **577**, 60 (2020).
10. F. Mohn, L. Gross, N. Moll, G. Meyer, *Nat. Nanotechnol.* **7**, 227 (2012).
11. T. Fukuma, Y. Ueda, S. Yoshioka, H. Asakawa, *Phys. Rev. Lett.* **104**, 016101 (2010).

ACKNOWLEDGMENTS

We acknowledge support from JSPS KAKENHI grants JP 20H05849, JP 22H04496, JP 22H05448, and JP 22H01950.

10.1126/science.add0841

MATERIALS SCIENCE

Building perovskite solar cells that last

An inorganic halide perovskite solar cell architecture promises multiyear stability

By **Severin N. Habisreutinger**[1] *and* **Matthew O. Reese**[2]

At the heart of a solar cell sits an absorber layer that converts sunlight into electricity. Metal-halide perovskites (MHPs) are a new class of such absorber materials, which have exceptional optoelectronic properties and can be manufactured by using low-cost, scalable solution-processing or vapor-based deposition methods. Consequently, perovskite photovoltaics (PV) have progressed at an unprecedented pace and have gone from experimental breakthroughs to the cusp of commercialization within the last decade. Thus far, much focus has been aimed at improving the power-conversion efficiency of small-area cells—now approaching 26%, having already surpassed most other thin-film PV technologies (*1*). Efficiency is, however, only one of many requirements for a PV technology to succeed. Equally important is the capacity to retain performance over time, for up to 25+ years. On page 307 of this issue, Zhao *et al.* (*2*) tackle this challenge of long-term stability of MHP solar cells through accelerated stress testing.

In contrast to covalent semiconductors like silicon, the mixed bonding—both covalent and ionic—in MHPs is a double-edged sword. It underlies the easy fabrication and exceptional defect tolerance of MHPs, but it also allows for a high degree of ionic mobility. As a result, external driving forces, such as light and heat, can redistribute ionic species within the absorber and cause material degradation. In particular, the egress and loss of volatile species have been identified as a major degradation mechanism in MHPs (*3*, *4*). This has raised the question of whether MHPs are a practical technology given the operational life span required of PV technologies.

Recently, various new approaches have been demonstrated to stabilize the MHP absorber by coating its surface with a thin capping layer (*5*–*7*). Zhao *et al.* took an extra step by designing a device architecture that relies exclusively on inorganic materials to improve resilience to thermal deformation and degradation. Beyond hardening the contact materials and electrodes, they also chose an inorganic cesium lead triiodide ($CsPbI_3$) MHP absorber. Although this leads to a lower initial efficiency, the absence of organic cations reduces possible thermal degradation processes. The authors also expanded on previous work using an inorganic Ruddlesden–Popper perovskite as the passivation layer on top of the absorber (*8*–*10*). This layer has two effects: It reduces nonradiative surface recombination of photogenerated charge carriers, thus improving the photovoltage (*11*); and it suppresses the egress of iodide ions from the absorber, thus preserving the integrity of the device stack. The importance of this strategy to make MHP solar cells more resilient is demonstrated through long-term device operation under continuous light exposure at several elevated temperatures, using the protocols developed by the International Summit on Organic Photovoltaic Stability (*12*, *13*).

Guaranteeing minimal losses under decades-long operation is an important factor for the "bankability" of a PV technology. This is a particular challenge for MHP solar cells, which simply have not been around long enough to validate long-term projections for real-world performance. Because of this, accelerated stress tests have been used to simulate longer periods of operation through more severe stressing conditions with acceleration factors. Zhao *et al.* used elevated temperatures beyond typical operational temperatures as acceleration factors, allowing them to determine an activation energy for the dominant degradation process.

Zhao *et al.* used an encapsulation approach for their devices to prevent extrinsic factors such as moisture and oxygen from playing a dominant role in device degradation. They subsequently tracked the performance of their encapsulated devices under operation for several months and noted that even for devices stressed at the highest temperature, it took almost 3 months for the efficiency to drop to 80% of the initial performance. With the calculated activation energy, the authors were able to translate the achieved stability at 110°C to 5+ years of continuous operation at

[1]Oxford PV, Yarnton, Kidlington, UK. [2]National Renewable Energy Laboratory (NREL), Golden, CO, USA. Email: severin.habisreutinger@oxfordpv.com; matthew.reese@nrel.gov

surface. This information is valuable for understanding the elementary processes involved in the hydrogen evolution reaction, by explaining the path taken by the protons going from the water molecule network to the substrate. Furthermore, Tian *et al.* clarified that the abundance ratio of Eigen and Zundel cations differs between platinum and gold substrates. Because platinum is a commonly used electrode material, Tian *et al.* highlight the importance of understanding the substrate dependence of the hydrogen evolution mechanism, including how proton transfer to the substrate may be affected by factors such as the lattice constant and surface potential of the substrate.

Tian *et al.* provide a direct observation of excess protons in the water molecule network. In addition to information on atomic positions obtained with AFM, the quantitative measurements of local charge distributions will also be useful for further analyzing more complex systems, such as hydrogen bonding networks that contain injected heterogeneous ions. Measuring the electrostatic force with AFM allows visualization of local charges within a single organic molecule (*10*). This approach will help determine the directionality of hydrogen bonds and identify ionic species.

High-resolution experiments have been conducted in ultrahigh vacuum at low temperature, whereas hydration structures have been observed at the solid-liquid interface with AFM at room temperature (*11*). The hydration structure is visualized as a time-averaged image of dynamically moving water molecules. It is important to conduct research that bridges the gap between static but high-resolution observations in ultrahigh vacuum and time-averaged observations in liquid that are close to more realistic conditions. In the future, real-time observation of proton transfers in the water adsorption layers in relation to electrochemical reactions will be a key technology relevant to various fields. ■

REFERENCES AND NOTES

1. Y. Tian *et al.*, *Science* **377**, 315 (2022).
2. T. Bartels-Rausch *et al.*, *Rev. Mod. Phys.* **84**, 885 (2012).
3. G. Binnig, C. F. Quate, C. Gerber, *Phys. Rev. Lett.* **56**, 930 (1986).
4. L. Gross, F. Mohn, N. Moll, P. Liljeroth, G. Meyer, *Science* **325**, 1110 (2009).
5. L. Gross *et al.*, *Angew. Chem. Int. Ed.* **57**, 3888 (2018).
6. A. Shiotari, Y. Sugimoto, *Nat. Commun.* **8**, 14313 (2017).
7. J. Peng *et al.*, *Nat. Commun.* **9**, 122 (2018).
8. A. Shiotari, Y. Sugimoto, H. Kamio, *Phys. Rev. Mater.* **3**, 093001(R) (2019).
9. R. Ma *et al.*, *Nature* **577**, 60 (2020).
10. F. Mohn, L. Gross, N. Moll, G. Meyer, *Nat. Nanotechnol.* **7**, 227 (2012).
11. T. Fukuma, Y. Ueda, S. Yoshioka, H. Asakawa, *Phys. Rev. Lett.* **104**, 016101 (2010).

ACKNOWLEDGMENTS

We acknowledge support from JSPS KAKENHI grants JP 20H05849, JP 22H04496, JP 22H05448, and JP 22H01950.

10.1126/science.add0841

MATERIALS SCIENCE

Building perovskite solar cells that last

An inorganic halide perovskite solar cell architecture promises multiyear stability

By **Severin N. Habisreutinger**[1] *and* **Matthew O. Reese**[2]

At the heart of a solar cell sits an absorber layer that converts sunlight into electricity. Metal-halide perovskites (MHPs) are a new class of such absorber materials, which have exceptional optoelectronic properties and can be manufactured by using low-cost, scalable solution-processing or vapor-based deposition methods. Consequently, perovskite photovoltaics (PV) have progressed at an unprecedented pace and have gone from experimental breakthroughs to the cusp of commercialization within the last decade. Thus far, much focus has been aimed at improving the power-conversion efficiency of small-area cells—now approaching 26%, having already surpassed most other thin-film PV technologies (*1*). Efficiency is, however, only one of many requirements for a PV technology to succeed. Equally important is the capacity to retain performance over time, for up to 25+ years. On page 307 of this issue, Zhao *et al.* (*2*) tackle this challenge of long-term stability of MHP solar cells through accelerated stress testing.

In contrast to covalent semiconductors like silicon, the mixed bonding—both covalent and ionic—in MHPs is a double-edged sword. It underlies the easy fabrication and exceptional defect tolerance of MHPs, but it also allows for a high degree of ionic mobility. As a result, external driving forces, such as light and heat, can redistribute ionic species within the absorber and cause material degradation. In particular, the egress and loss of volatile species have been identified as a major degradation mechanism in MHPs (*3*, *4*). This has raised the question of whether MHPs are a practical technology given the operational life span required of PV technologies.

Recently, various new approaches have been demonstrated to stabilize the MHP absorber by coating its surface with a thin capping layer (*5*–*7*). Zhao *et al.* took an extra step by designing a device architecture that relies exclusively on inorganic materials to improve resilience to thermal deformation and degradation. Beyond hardening the contact materials and electrodes, they also chose an inorganic cesium lead triiodide ($CsPbI_3$) MHP absorber. Although this leads to a lower initial efficiency, the absence of organic cations reduces possible thermal degradation processes. The authors also expanded on previous work using an inorganic Ruddlesden–Popper perovskite as the passivation layer on top of the absorber (*8*–*10*). This layer has two effects: It reduces nonradiative surface recombination of photogenerated charge carriers, thus improving the photovoltage (*11*); and it suppresses the egress of iodide ions from the absorber, thus preserving the integrity of the device stack. The importance of this strategy to make MHP solar cells more resilient is demonstrated through long-term device operation under continuous light exposure at several elevated temperatures, using the protocols developed by the International Summit on Organic Photovoltaic Stability (*12*, *13*).

Guaranteeing minimal losses under decades-long operation is an important factor for the "bankability" of a PV technology. This is a particular challenge for MHP solar cells, which simply have not been around long enough to validate long-term projections for real-world performance. Because of this, accelerated stress tests have been used to simulate longer periods of operation through more severe stressing conditions with acceleration factors. Zhao *et al.* used elevated temperatures beyond typical operational temperatures as acceleration factors, allowing them to determine an activation energy for the dominant degradation process.

Zhao *et al.* used an encapsulation approach for their devices to prevent extrinsic factors such as moisture and oxygen from playing a dominant role in device degradation. They subsequently tracked the performance of their encapsulated devices under operation for several months and noted that even for devices stressed at the highest temperature, it took almost 3 months for the efficiency to drop to 80% of the initial performance. With the calculated activation energy, the authors were able to translate the achieved stability at 110°C to 5+ years of continuous operation at

[1]Oxford PV, Yarnton, Kidlington, UK. [2]National Renewable Energy Laboratory (NREL), Golden, CO, USA. Email: severin.habisreutinger@oxfordpv.com; matthew.reese@nrel.gov

35°C. It is worth highlighting that real-world operating conditions vary by geographic location and mounting conditions. For example, a module may occasionally experience 85°C in Saudi Arabia, with a time-averaged temperature of up to 55°C, whereas the climatic conditions will be much more benign in temperate regions such as Germany, where there are fewer extreme temperature spikes and an average temperature well below 35°C (*14*).

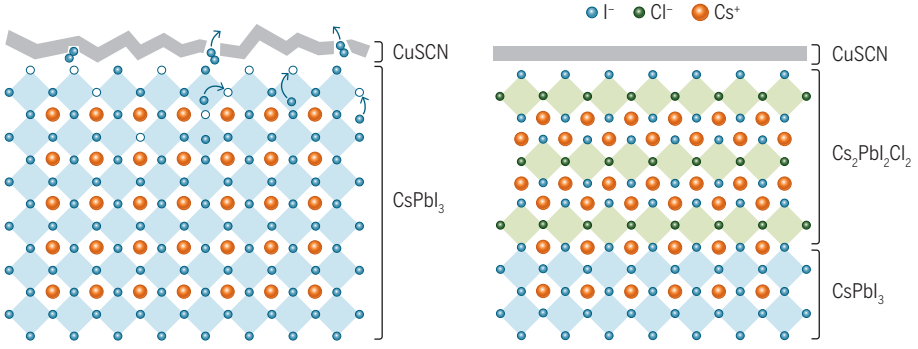
With this study, Zhao *et al.* demonstrated high intrinsic stability of a hardened device to temperature and light, indicating that material choice and device engineering can provide a pathway to minimizing intrinsic performance losses to unavoidable stressors. Although this device stack has a raw performance much worse than the current state of the art, its long-term performance approaches levels necessary for real-world bankability.

However, acceleration factors are a tricky business and are not easily translatable for predicting real-world performance. In this case, temperature-activated iodine egress was identified as the dominant degradation mechanism with a single temperature-dependent acceleration factor. Other material systems or device architectures are likely to have different activation temperatures and possibly multiple interdependent degradation mechanisms, resulting in more complex degradation dynamics. When designing these accelerated testing protocols, it is important to avoid triggering degradation mechanisms that are not present in real-world operating conditions—baking a pie at 2000°C for 6 minutes is not the same as baking it at 200°C for an hour. This limits the viable temperature range and hence magnitude of the acceleration factor. It is also not clear yet whether further degradation pathways might be triggered by additional stressors such as diurnal and temperature cycling, partial shading of modules, and potential-induced degradation from high voltages. Thus, cells and mini-modules must be field-tested outdoors to identify the dominant real-world degradation pathways and correlate them to accelerated lab tests.

It is also critical to highlight that although Zhao *et al.* demonstrated impressive stability in their devices when encapsulated to keep out oxygen and moisture, their unencapsulated devices degraded 1000 times faster under the same testing conditions. Robust encapsulation strategies from the thin-film PV community exist, but it is arguably risky to solely rely on encapsulation for long-term stability, as packages can fail. The strength of Zhao *et al.* is the formulation of a cogent hypothesis, which, through careful testing and validation, generated tangible insights. The challenge for future studies will be to apply these insights to device architectures that pair enhanced intrinsic stability with the superior efficiencies of the state of the art. ■

An inorganic capping layer for an inorganic perovskite solar cell

A dominant degradation process for the three-dimensional (3D) $CsPbI_3$ perovskite is the egress of iodine, which can corrode the contacts of the solar cell (left). Zhao *et al.* coated the 3D absorber with an inorganic 2D perovskite layer ($Cs_2PbI_2Cl_2$), before capping it off with copper(I) thiocyanate (CuSCN). The design inhibits the egress of iodine and minimizes defects at the interface, thus enhancing the stability and longevity of the device.



REFERENCES AND NOTES

1. M. A. Green *et al.*, *Prog. Photovolt. Res. Appl.* **30**, 3 (2022).
2. X. Zhao *et al.*, *Science* **377**, 307 (2022).
3. T. Leijtens *et al.*, *Adv. Energy Mater.* **5**, 1500963 (2015).
4. C. C. Boyd *et al.*, *Chem. Rev.* **119**, 3418 (2019).
5. X. Li *et al.*, *Science* **375**, 434 (2022).
6. Z. Li *et al.*, *Science* **376**, 416 (2022).
7. R. Azmi *et al.*, *Science* **376**, 73 (2022).
8. J. Li *et al.*, *J. Am. Chem. Soc.* **140**, 11085 (2018).
9. H. Wang *et al.*, *Nano Energy* **89**, 106411 (2021).
10. Z. Li *et al.*, *J. Phys. Chem. Lett.* **11**, 4138 (2020).
11. D. L. McGott *et al.*, *Joule* **5**, 1057 (2021).
12. M. O. Reese *et al.*, *Sol. Energy Mater. Sol. Cells* **95**, 1253 (2011).
13. M. V. Khenkin *et al.*, *Nat. Energy* **5**, 35 (2020).
14. M. D. Kempe, D. Panchagade, M. O. Reese, A. A. Dameron, *Prog. Photovolt. Res. Appl.* **23**, 570 (2015).

ACKNOWLEDGMENTS

This work was produced by Alliance for Sustainable Energy, LLC under contract no. DE-AC36-08GO28308 with the US Department of Energy (DOE). Publisher acknowledges the US Government license to provide public access under the DOE Public Access Plan. Funding was provided by the US DOE Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office.

10.1126/science.abq7013

AGING

Mosaic loss of Y chromosome during aging

Spread of Y chromosome aneuploidy in myeloid cells with age promotes cardiac fibrosis

By **Andreas Zeiher**[1,2] *and* **Thomas Braun**[2,3]

During their lifetimes, organisms acquire somatic mutations in individual cells caused by genomic instability, endogenous DNA replication errors, or exposure to mutagens (*1*). When mutations occur in somatic stem cells, the mutation "spreads" in a mosaic manner, appearing in the progeny of mutated stem cells but not in cells from nonmutated stem cells. Mosaic loss of Y chromosome (mLOY) has been observed in peripheral leukocytes of aging men (*2*), reaching 40% of leukocytes in individuals over 70 years (*3*). LOY increases over time and correlates with clonal expansion of myeloid cells (*4*). LOY also correlates with increased risk for mortality, cardiovascular events, and other age-associated disorders, but a causal relationship has not been established (*2*, *5*). On page 292 of this issue, Sano *et al.* (*6*) determine that mLOY in myeloid cells is a major risk factor for increased incidence of cardiovascular and fibrotic diseases during aging.

Mutations attained during adulthood become particularly threatening when they occur in somatic stem cells and result in the acquisition of competitive advantages (*7*). In solid organs, mutated cell clones will be primarily confined to the organ of origin, whereas clonal expansion of mutated stem cells within the hematopoietic system potentially reaches multiple organ systems. In particular, circulating cells of the myeloid lineage can traffic to solid organs to help maintain tissue homeostasis and repair. Mutations in hematopoietic cells

[1]Institute of Cardiovascular Regeneration, Goethe University, Frankfurt, Germany. [2]German Center for Cardiovascular Research DZHK, Partner Site Rhine-Main, Berlin, Germany. [3]Max Planck Institute for Heart and Lung Research, Department of Cardiac Development and Remodeling, Bad Nauheim, Germany.
Email: thomas.braun@mpi-bn.mpg.de

GRAPHIC: K. FRANKLIN/*SCIENCE*

35°C. It is worth highlighting that real-world operating conditions vary by geographic location and mounting conditions. For example, a module may occasionally experience 85°C in Saudi Arabia, with a time-averaged temperature of up to 55°C, whereas the climatic conditions will be much more benign in temperate regions such as Germany, where there are fewer extreme temperature spikes and an average temperature well below 35°C (*14*).

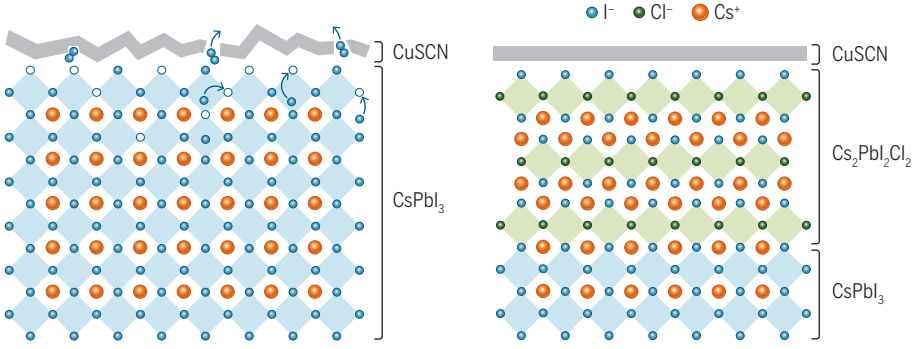
With this study, Zhao *et al.* demonstrated high intrinsic stability of a hardened device to temperature and light, indicating that material choice and device engineering can provide a pathway to minimizing intrinsic performance losses to unavoidable stressors. Although this device stack has a raw performance much worse than the current state of the art, its long-term performance approaches levels necessary for real-world bankability.

However, acceleration factors are a tricky business and are not easily translatable for predicting real-world performance. In this case, temperature-activated iodine egress was identified as the dominant degradation mechanism with a single temperature-dependent acceleration factor. Other material systems or device architectures are likely to have different activation temperatures and possibly multiple interdependent degradation mechanisms, resulting in more complex degradation dynamics. When designing these accelerated testing protocols, it is important to avoid triggering degradation mechanisms that are not present in real-world operating conditions—baking a pie at 2000°C for 6 minutes is not the same as baking it at 200°C for an hour. This limits the viable temperature range and hence magnitude of the acceleration factor. It is also not clear yet whether further degradation pathways might be triggered by additional stressors such as diurnal and temperature cycling, partial shading of modules, and potential-induced degradation from high voltages. Thus, cells and mini-modules must be field-tested outdoors to identify the dominant real-world degradation pathways and correlate them to accelerated lab tests.

It is also critical to highlight that although Zhao *et al.* demonstrated impressive stability in their devices when encapsulated to keep out oxygen and moisture, their unencapsulated devices degraded 1000 times faster under the same testing conditions. Robust encapsulation strategies from the thin-film PV community exist, but it is arguably risky to solely rely on encapsulation for long-term stability, as packages can fail. The strength of Zhao *et al.* is the formulation of a cogent hypothesis, which, through careful testing and validation, generated tangible insights. The challenge for future studies will be to apply these insights to device architectures that pair enhanced intrinsic stability with the superior efficiencies of the state of the art. ■

An inorganic capping layer for an inorganic perovskite solar cell

A dominant degradation process for the three-dimensional (3D) $CsPbI_3$ perovskite is the egress of iodine, which can corrode the contacts of the solar cell (left). Zhao *et al.* coated the 3D absorber with an inorganic 2D perovskite layer ($Cs_2PbI_2Cl_2$), before capping it off with copper(I) thiocyanate (CuSCN). The design inhibits the egress of iodine and minimizes defects at the interface, thus enhancing the stability and longevity of the device.



REFERENCES AND NOTES

1. M. A. Green *et al.*, *Prog. Photovolt. Res. Appl.* **30**, 3 (2022).
2. X. Zhao *et al.*, *Science* **377**, 307 (2022).
3. T. Leijtens *et al.*, *Adv. Energy Mater.* **5**, 1500963 (2015).
4. C. C. Boyd *et al.*, *Chem. Rev.* **119**, 3418 (2019).
5. X. Li *et al.*, *Science* **375**, 434 (2022).
6. Z. Li *et al.*, *Science* **376**, 416 (2022).
7. R. Azmi *et al.*, *Science* **376**, 73 (2022).
8. J. Li *et al.*, *J. Am. Chem. Soc.* **140**, 11085 (2018).
9. H. Wang *et al.*, *Nano Energy* **89**, 106411 (2021).
10. Z. Li *et al.*, *J. Phys. Chem. Lett.* **11**, 4138 (2020).
11. D. L. McGott *et al.*, *Joule* **5**, 1057 (2021).
12. M. O. Reese *et al.*, *Sol. Energy Mater. Sol. Cells* **95**, 1253 (2011).
13. M. V. Khenkin *et al.*, *Nat. Energy* **5**, 35 (2020).
14. M. D. Kempe, D. Panchagade, M. O. Reese, A. A. Dameron, *Prog. Photovolt. Res. Appl.* **23**, 570 (2015).

ACKNOWLEDGMENTS

This work was produced by Alliance for Sustainable Energy, LLC under contract no. DE-AC36-08GO28308 with the US Department of Energy (DOE). Publisher acknowledges the US Government license to provide public access under the DOE Public Access Plan. Funding was provided by the US DOE Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office.

10.1126/science.abq7013

AGING

Mosaic loss of Y chromosome during aging

Spread of Y chromosome aneuploidy in myeloid cells with age promotes cardiac fibrosis

By **Andreas Zeiher**[1,2] *and* **Thomas Braun**[2,3]

During their lifetimes, organisms acquire somatic mutations in individual cells caused by genomic instability, endogenous DNA replication errors, or exposure to mutagens (*1*). When mutations occur in somatic stem cells, the mutation "spreads" in a mosaic manner, appearing in the progeny of mutated stem cells but not in cells from nonmutated stem cells. Mosaic loss of Y chromosome (mLOY) has been observed in peripheral leukocytes of aging men (*2*), reaching 40% of leukocytes in individuals over 70 years (*3*). LOY increases over time and correlates with clonal expansion of myeloid cells (*4*). LOY also correlates with increased risk for mortality, cardiovascular events, and other age-associated disorders, but a causal relationship has not been established (*2*, *5*). On page 292 of this issue, Sano *et al.* (*6*) determine that mLOY in myeloid cells is a major risk factor for increased incidence of cardiovascular and fibrotic diseases during aging.

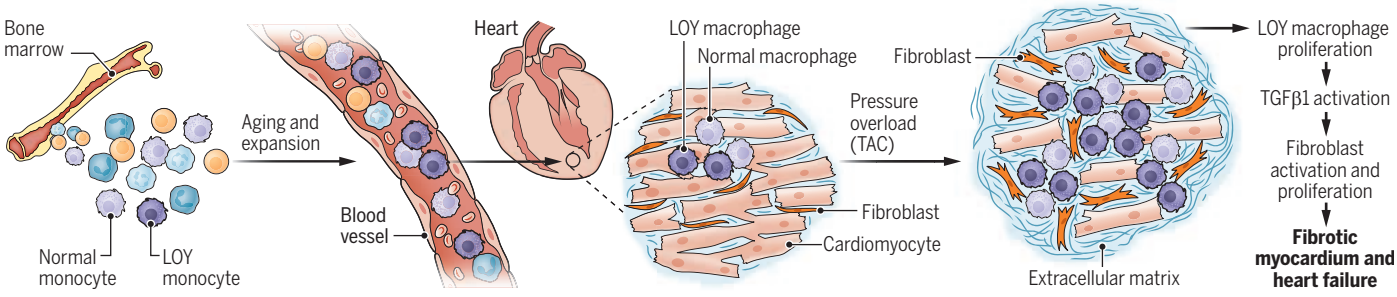
Mutations attained during adulthood become particularly threatening when they occur in somatic stem cells and result in the acquisition of competitive advantages (*7*). In solid organs, mutated cell clones will be primarily confined to the organ of origin, whereas clonal expansion of mutated stem cells within the hematopoietic system potentially reaches multiple organ systems. In particular, circulating cells of the myeloid lineage can traffic to solid organs to help maintain tissue homeostasis and repair. Mutations in hematopoietic cells

[1]Institute of Cardiovascular Regeneration, Goethe University, Frankfurt, Germany. [2]German Center for Cardiovascular Research DZHK, Partner Site Rhine-Main, Berlin, Germany. [3]Max Planck Institute for Heart and Lung Research, Department of Cardiac Development and Remodeling, Bad Nauheim, Germany.
Email: thomas.braun@mpi-bn.mpg.de

GRAPHIC: K. FRANKLIN/*SCIENCE*

Mosaic loss of Y chromosome in myeloid cells induces cardiac fibrosis

Bone marrow–derived myeloid cells exhibit loss of Y chromosome (LOY) during aging. They expand and colonize different organs, causing organ fibrosis, particularly in hearts subjected to pressure overload by transaortic constriction (TAC). LOY is enriched in profibrotic macrophages with high transforming growth factor β1 (TGFβ1) activity, triggering a TGFβ1-dependent feed-forward signaling loop that results in fibroblast proliferation and increased extracellular matrix deposition in mice.



may lead to clonal expansion of peripheral blood cells without necessarily progressing to myeloid neoplasms, a condition called clonal hematopoiesis of indeterminate potential (CHIP) (*8*). CHIP increases with age, approaching an incidence of ~12% in individuals over 70 years of age, with mutant clones accounting for ~20% of peripheral blood mononuclear cells on average (*9*). Notably, individuals with CHIP not only exhibit increased risk for hematological cancer but also show higher incidence of cardiovascular diseases, which is the main cause of increased CHIP-associated mortality (*8*). The most prevalent mutated genes in CHIP are DNA methyltransferase 3A (*DNMT3A*) and tet methylcytosine dioxygenase 2 (*TET2*), which are linked with increased expression of inflammatory genes in myeloid cells that instigate atherosclerosis, impaired heart function, and cardiac fibrosis upon their recruitment into cardiovascular tissues (*10*).

mLOY shows some similarities to CHIP, such as prevalence in the myeloid lineage, increased incidence during aging, and association with chronic diseases. Mechanistic studies for mLOY were completely lacking, preventing the establishment of a causal relationship between LOY and cardiovascular disorders. Sano *et al.* generated a mouse model of mLOY in which the centromere of the Y chromosome was deleted in hematopoietic stem cells, which leads to specific loss of this chromosome. No obvious hematological abnormalities were observed in mLOY mice, in which 49 to 81% of white blood cells are devoid of the Y chromosome, comparable to men with mLOY. However, mLOY in mice led to cardiac fibrosis during aging, particularly after left ventricular pressure overload due to transaortic constriction. LOY macrophages, recruited to the heart, showed aberrant profibrotic differentiation and higher activity of the profibrotic cytokine transforming growth factor β1 (TGFβ1). Unexpectedly, the majority of TGFβ1 signaling in mLOY hearts after transaortic constriction was detected in other cell types within the myocardium, suggesting that aberrantly polarized mLOY macrophages, which release TGFβ1, trigger a feed-forward signaling loop that activates fibroblasts or other cells in the myocardium, driving fibrosis (see the figure). Furthermore, a prospective analysis of mLOY in 223,173 men using data from the UK Biobank revealed that during an average follow-up time of 11.5 years, men with LOY in >40% of leukocytes exhibited a 31% increased risk of death from cardiovascular disorders, including hypertensive heart disease and congestive heart failure, as well as peripheral vascular disorders such as aortic aneurysm and dissection.

The Y chromosome has been mainly viewed as a "genetic wasteland," whose dominant role is to determine gonadal sex but little else. In comparison to other chromosomes, the Y chromosome is gene poor, and repetitive elements make up more than 50% of its genetic material. Nevertheless, the Y chromosome has also been considered as a potential "blueprint for men's health" [(*11*), p. 1181]. Indeed, several unexpected links between the Y chromosome, immune system, and complex polygenic traits have been discovered, suggesting an influence of the Y chromosome on immune and inflammatory responses in men (*11*). The study of Sano *et al.* reinforces this view and uncovers a crucial function of the Y chromosome in maintaining a healthy innate immune system, but further research is required to elucidate the mechanisms. Important mechanistic questions that remain include how the Y chromosome is lost during aging. Whether mLOY is a particular feature of the hematopoietic system or also happens in solid organs requires further investigation. It is also unclear what drives the TGFβ1-dependent feed-forward loop between mLOY macrophages and cells in the myocardium during aging.

Co-occurrence of CHIP and mLOY has been reported in monocytes from a small cohort of 24 healthy men, suggesting that both abnormalities can coexist (*12*). It will be important to determine the frequency of separate or combined occurrence and potential consequences for disease manifestation and prognosis. Does the concomitant presence of mLOY and CHIP create new effects and further alter the function of bone marrow–derived cells? For example, CHIP mutations may promote expansion of LOY cells. In addition, mLOY seems to primarily favor fibrotic pathologies, whereas CHIP caused by *TET2* or *DNMT3A* mutations enhances inflammatory processes, eventually culminating in organ fibrosis (*10*). Perhaps the coexistence of mLOY and CHIP fuels a vicious cycle leading to organ fibrosis during aging. Removal of defective myeloid cells that either promote inflammation (CHIP) or promote fibrosis (mLOY) is an attractive therapeutic option to prevent or stall chronic disease processes that lead to progressive organ failure. Thorough characterization of mLOY cells may help to identify a target, or targets, that can be exploited to ablate such cells. ■

GRAPHIC: K. HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. F. Manders, R. van Boxtel, S. Middelkamp, *Front. Aging* **2**, 802407 (2021).
2. L. A. Forsberg *et al.*, *Nat. Genet.* **46**, 624 (2014).
3. D. J. Thompson *et al.*, *Nature* **575**, 652 (2019).
4. M. Danielsson *et al.*, *Eur. J. Hum. Genet.* **28**, 349 (2020).
5. S. Haitjema *et al.*, *Circ. Cardiovasc. Genet.* **10**, e001544 (2017).
6. S. Sano *et al.*, *Science* **377**, 292 (2022).
7. S. Olafsson, C. A. Anderson, *Trends Genet.* **37**, 872 (2021).
8. D. P. Steensma *et al.*, *Blood* **126**, 9 (2015).
9. S. Jaiswal, P. Libby, *Nat. Rev. Cardiol.* **17**, 137 (2020).
10. S. Sano *et al.*, *Circ. Res.* **123**, 335 (2018).
11. A. A. Maan *et al.*, *Eur. J. Hum. Genet.* **25**, 1181 (2017).
12. V. Ljungström *et al.*, *Leukemia* **36**, 889 (2022).

ACKNOWLEDGMENTS

A.Z. and T.B. are supported by the Cardiopulmonary Institute funded by the German Research Foundation (Exc2026). T.B. is supported by the German Research Foundation, project no. 268555672–SFB 1213, projects A02 and B02.

10.1126/science.add0839

INSIGHTS

POLICY FORUM

ILLUSTRATION: DAVIDE BONAZZI/SALZMANART

SCIENCE INFRASTRUCTURE

Playing catch-up in building an open research commons

As efforts advance around the globe, the US falls behind

By **Philip E. Bourne[1], Vivien Bonazzi[2], Amy Brand[3], Bonnie Carroll[4], Ian Foster[5], Ramanathan V. Guha[6], Robert Hanisch[7], Sallie Ann Keller[8], Mary Lee Kennedy[9], Christine Kirkpatrick[10], Barend Mons[11,12], Sarah M. Nusser[13], Michael Stebbins[14], George Strawn[15], Alex Szalay[16]**

Many aspects of the research enterprise are rapidly changing to be more open, accessible, and supportive of rapid-response investigations (e.g., understanding COVID-19) and large cross-national research that addresses complex challenges (e.g., supply chain issues). Around the globe, there have been aggressive responses to the need for a unified open research commons (ORC)—an interoperable collection of data and compute resources within both the public and private sectors that is easy to use and accessible to all. Many nations are positioning themselves to be scientifically competitive in the years to come. But the US is falling behind in the accessibility and connectedness of its research computing and data infrastructure (*1*), compromising competitiveness and leadership and limiting global science that could benefit from US contributions. The challenge is more cultural and institutional than technical and demands immediate and sustained leadership and support, starting with policy-makers and research funders.

The value of cooperation around technology and data broadly is beyond question. For example, shared governance, shared infrastructure, and agreements on standards that permit a shared system to operate underpin the ability of the North American electrical grid to direct electricity to where it is needed

and of the Cirrus banking network to deliver money from one's bank to almost anywhere in the world. The sum of the parts is made greater than the whole through cooperation across the enterprise.

Similar coordination in the research enterprise could pay enormous dividends. We now have vast amounts of publicly available research data, but to fully leverage the potential power of these data beyond individual and often heroic efforts, these data need to be identified, made interoperable, and aligned so that they can be broadly used by the scientific community. For example, often data on disparate topics—such as a county's homelessness rates, average income, neighborhood food and health resources, air pollution, flood risk, predicted water resources, and predicted average temperature—are spread across a range of locations on the web, infrastructures, and management regimes. If these data were integrated (brought together based on common data elements in each dataset), we could use the data for powerful analyses, like identifying locations with high homeless populations that are also likely to be hit hardest by floods, droughts, or heat waves or places with poor cardiac health that also have high or increasing particulate matter 2.5 (PM 2.5) pollution, which could lead to more heart attacks. Support by policy-makers and funders who are driving the development of research infrastructure can facilitate such work, similar to the urgent cooperation we see among scientists during times of dire need, such as the COVID-19 pandemic, the threat of war, and the disruption to the global economy.

In principle, this should be possible in the US, which has a vibrant research ecosystem with no lack of computation and data resources. But establishing an ORC is less a technical challenge than a cultural and institutional one that requires policy leadership and a sustained commitment, both of which have been lacking. Talk to any of the burgeoning number of data scientists and they will likely tell you about, in accessing data that are already publicly available, all the time they must spend learning a variety of esoteric compute systems and figuring out where to get data and what value the data have before they even begin the real work of analysis and discovery. Common application programming interfaces, new metadata and data standards, workflows (*2*), dashboards, and evaluation of progress as these evolve would enable data scientists to better solve the problems facing society. Beyond the technical aspects, there are also social and cultural aspects of appropriate recognition that data are first-class research objects that should be cited and acknowledged in the same way that publications are today.

In recognition of these challenges and opportunities, in 2013, the White House Office of Science and Technology Policy (OSTP) issued a directive to federal agencies to improve access to the results of scientific research outcomes. The intentions of the administration were made clear: "digitally formatted scientific data resulting from unclassified research supported wholly or in part by federal funding should be stored and publicly accessible to search, retrieve, and analyze." This US initiative was an impetus to the global push to open government, open science, and open data.

Countries around the world have since forged ahead. A wide range of recent efforts reflect substantial recognition, in other countries and regions, of the important role played by ORCs, for example, the European Open Science Cloud (EOSC), the CS3MESH4EOSC Science Mesh, the China Science and Technology (CST) Cloud, the African Open Science Platform, the South African National Integrated Cyber Infrastructure System, the Malaysia Open Science Platform, the Global Open Science Cloud (GOSC) (funded mainly by China and organized through the International Science Council's Committee on Data), the Australian Research Data Commons (ARDC) Nectar Research Cloud, the Digital Research Alliance of Canada (formerly known as the New Digital Research Infrastructure Organization), and the Arab States Research and Education Network.

Although all are quite recent initiatives, the CST Cloud already serves more than 1 million researchers. ARDC is supporting studies of bushfires, which have plagued Australia in recent years, and is working with all 43 universities across Australia to build a nationally agreed-upon network for research data. The EOSC-related projects and the European member states participate with considerable additional investments governed through the EOSC association spanning the full scope of scientific disciplines and provide support to high-performance computing centers, massive databases, and the software tools required to use them. As a result, the EOSC is pivotal in supporting research challenges such as climate change, space weather, seismology, bioinformatics, disaster mitigation, toxicology, and radio astronomy. Notwithstanding such early signs of progress, systematic evaluation of these developing projects is needed to assess their full value.

Critically, the governance models for many of these programs seek to enable equitable access to research capacity, consonant with a major policy goal of the current US administration. For example, substantial effort, as stated in the EOSC declaration, went into defining rules of participation, processes for governance, and allocation of resources.

FRAGMENTED AND FALLING BEHIND

As these collaborative initiatives advance around the globe, anecdotally, progress in establishing an ORC has seemed to fall behind in the US because of a vacuum of leadership, focus, and coordination. Without a coherent national strategy, US scientists have found it harder to participate in the broader global pursuit of ORCs, eroding US competitiveness.

The major US research funding agencies largely tend to pursue independent initiatives and rarely work together on shared infrastructure, even as the problems faced by society span agencies. Aside from a few shared resources such the Extreme Science and Engineering Discovery Environment (XSEDE) of the National Science Foundation (NSF), with its focus on the use of high-performance computing, and the Open Science Grid (OSG) Consortium, with shared compute and data resources (mostly NSF sponsored), we assert that US research computational and data infrastructure remains fragmented, inefficient, and uncompetitive with ORCs that are emerging globally.

Funding would seem to be diminishing, as in the case of XSEDE's follow-on program, Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS), which is resourced to provide roughly half of the funds of XSEDE despite increased usage and reliance of US scientists on advanced computing. The OSG Consortium, by contrast, is more robust, with a distributed funding model, yet has a single administrative council and software support services.

At the same time, there is little equity to the resources that exist. Access tends to favor the major research-intensive universities where substantial expertise and local compute access already exist. More can be achieved by pooling publicly available data and compute resources and enabling a larger and more diverse group of researchers, who

[1]School of Data Science, University of Virginia, Charlottesville, VA, USA. [2]Deloitte, Washington, DC, USA. [3]MIT Press, Cambridge, MA, USA. [4]Information International Associates, Oak Ridge, TN, USA. [5]Argonne National Laboratory, Lemont, IL, USA. [6]Google, Mountain View, CA, USA. [7]Office of Data and Informatics, National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA. [8]Biocomplexity Institute, University of Virginia, Charlottesville, VA, USA. [9]Association of Research Libraries, Washington, DC, USA. [10]San Diego Supercomputer Center, San Diego, CA, USA. [11]Human Genetics Department, Leiden University, Leiden, Netherlands. [12]Leiden Academic Center for Drug Research, Leiden University, Leiden, Netherlands. [13]Department of Statistics, Iowa State University, Ames, IA, USA. [14]Science Advisors LLC, Alexandria, VA, USA. [15]Independent researcher, Vienna, VA, USA. [16]Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA. Email: peb6a@virginia.edus

are from underrepresented communities and found in universities and elsewhere, to easily access those resources to address society's most pressing problems.

The tech industry is working to provide some integration of and access to data, but it is often not done with the proper research context and is driven by profit potential, not scientific need. The research community should embrace US industry potential where mutual benefit can be found, thus partnering and building trust rather than being dependent on industry. Yet public-private partnerships are few and limited by government rules, leading to separate organizations such as the Foundation for the NIH (FNIH), which facilitates joint initiatives between the National Institutes of Health (NIH) and the private sector (e.g., the Accelerating Medicines Partnership) in ways not allowed by a federal agency alone. Hence, the interdisciplinary knowledge, innovation, and underlying shared data and compute resources needed to solve global challenges are usually lacking.

The lack of a unified ORC leaves the US following along in global efforts (e.g., GOSC) with no formal executive representation in international ORC-focused initiatives such as at the Open Science Clouds and Commons Executives' Roundtable, where executives of many of the international initiatives collaborate in creating an interconnected global research cloud infrastructure built from the individual research commons developed by other nations.

We are encouraged by recent efforts to improve access to computation and data infrastructure, for example, in the National Artificial Intelligence Initiative. However, this represents a partial and siloed effort to serve specific communities that are largely focused on machine learning and artificial intelligence (AI) applications. Although important, there is a larger research ecosystem that needs access to data and computational resources.

Data, in particular, need to be made AI ready, and be preserved for substantial periods of time in order to support reproducibility to ensure the integrity of the scholarly research record. Such preservation is critical because data often cost so much to produce as to be irreplaceable and/or because the aggregation of such data may lead to new insights at some indeterminate future time. But the long-term beneficiaries may be very different from the producers or initial custodians, which adds to the challenge and needs to be part of a national ORC concept. This requires professional data stewardship, a skill that is strongly developed in the various regional initiatives around the globe and exemplified by the International Society for Biocuration.

ALIGNING INCENTIVES

It is imperative that all stakeholders together create a more seamlessly connected and accessible data and compute infrastructure. What is needed is commitment from the US Congress and the administration so that parties can come together to chart the future and address the fragmented nature of the US research computing and data enterprise by establishing an ORC. Building on prior conceptualizations (*3*, *4*), the ORC should span federal, state, and local government agencies and computing facilities, including national laboratories, public and private clouds, and institutions. The ORC should be designed to replace the largely siloed, individually controlled data and compute resources present in the US today, a situation that limits discoverability, access, innovation, and collaboration.

This will require shared governance, trust, common standards, shared infrastructure, and, crucially, a champion, such as the OSTP.

"...establishing an open research commons... requires policy leadership and a sustained commitment..."

A key to success is the incentive to create a unified system. Scientists are not yet presented with the adequate incentives. Mandates from funders, such as data-sharing policies, help, but there are not enough definitions of requirements and rewards for complying or, indeed, a unification of what is expected of researchers regardless of the source of their research funding. What is lacking are the coordination and incentives to establish shared data and compute resources, governance, policies, and procedures across the US research enterprise to maximize accessibility and equity and train a computation- and data-savvy workforce.

In addition to the need to be competitive is the need to contribute to the global ecosystem, such that the will to collaborate is complemented by the ORC infrastructure to make it productive to do so. Efforts in the US must thus absolutely learn from and engage with others abroad, to adopt and adapt rather than start from scratch and reinvent what is already known. Efforts must reach out across sectors. Though commercial entities must be part of the ecosystem, they alone cannot be expected to adequately develop an ORC that serves the public good and needs of scientists, particularly as a long-term commitment. An ORC would thus provide a focal point for creating and nurturing public-private partnerships by adopting shared research data policies, standards, and practices, as well as the technical infrastructure to reduce the risk of absent, inoperable, and lost knowledge at the time of need.

Much of the hardware, software, and knowledge needed to make the ORC a reality largely exists. Thus, the cost may not be overly daunting compared with the gains in productivity, insight, economic competitiveness, and security that may emerge. Although the exact return on investment is hard to estimate (more explicit collaborative research is needed here), an increase in research productivity at even the 10% level would likely far outweigh the cost of establishing the ORC.

Recent developments in the US point to expected benefits of a unified and coordinated approach. Notable developments are the NSF's Research Coordination Network and the NSF program of investment known as the Research Data Ecosystem: A National Resource for Reproducible, Robust, and Transparent Social Science Research in the 21st Century. But much more is needed.

Congress and the administration must take action now. That action should be defined by the OSTP and include mandates to all federally funded science and technology agencies: Mandates for cooperation across agencies, like that currently being attempted by the Global Biodata Coalition, albeit more focused on sustainability, that lead to shared compute infrastructure and, if not shared, governed by a set of rules and standards that facilitate data exchange and reuse and provide a consistent interface to both humans and machines; and mandates that have the agencies working in closer cooperation with the private sector to realize the full potential of the US technology workforce without compromising competition. Only then will there be opportunities to maximize productivity and innovation to solve global problems. Other countries and regions are taking these same steps. It is time for the US to step up to the plate. ■

REFERENCES AND NOTES

1. I. Foster, D. Lopresti, B. Gropp, M. D. Hill, K. Schauman, arXiv:2104.06953 [cs.CY] (2021).
2. National Academies of Sciences, Engineering, and Medicine, *Automated Research Workflows for Accelerated Discovery: Closing the Knowledge Discovery Loop* (National Academies Press, 2022).
3. E. Ostrom, *Governing the Commons* (Cambridge Univ. Press, 1990).
4. B. M. Frischmann, M. J. Madison, K. J. Strandburg, Eds., *Governing Knowledge Commons* (Oxford Scholarship Online, 2014).

ACKNOWLEDGMENTS

This commentary draws upon discussions held at an event, "Towards a Unified US Open Research Commons," that was held on 2 August 2021 in Washington, DC, with support from the National Academies of Sciences, Engineering, and Medicine. The event was inspired in part by (*1*). These opinions, recommendations, findings, and conclusions do not necessarily reflect the views or policies of the National Institute of Standards and Technology (NIST) or the United States government.

10.1126/science.abo5947

BOOKS *et al.*

Cathedral mounds like this one in Litchfield National Park are constructed by hive-minded termites.

NEUROSCIENCE

An exploration of real, virtual, and possible minds

What are they for—who has them and why?

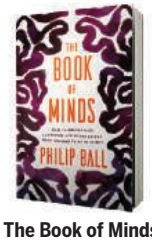
By **Alex Gomez-Marin**

With ambition and patience, in *The Book of Minds*, British science writer Philip Ball explores the parameters and functions of actual, virtual, and possible minds. The journey begins with humans and our fellow organisms on Earth (including plants and fungi) and ends with machine-based minds (artificial intelligence) and minds beyond our grasp (extraterrestrials and even God).

The Book of Minds addresses everything from intelligence and consciousness to agency and free will. In doing so, Ball risks biting off more than we can chew. And yet the book reads swiftly and smoothly. Organized into 10 generous chapters, the book often feels like the educated rambling of a passionate dilettante. Ball's take is wide and balanced, likely to please the demigods of mainstream academe while veiling valiant minority reports in its interstices.

What is a mind? According to Ball, "For an entity to have a mind, there must be something it is like to be that entity." Starting with human brains, he acknowledges the constitutive role of bodies, emotions, and the environment. He discusses socialization, language, and the evolution of intelligence, insisting on "the constructive faculty of mind." Ball then turns to consciousness, mentioning the work of the celebrities in this arena, from philosophers David Chalmers and Daniel Dennett to neuroscientists Antonio Damasio and Stanislas Dehaene, in a comprehensive but predictable chapter.

When it comes to other animals, Ball warns readers against conceiving of them as "dim-witted humans." Jakob von Uexküll's concept of "Umwelt" helps here, he insists; every organism experiences its own meaningful environment. From ravens to great apes, Ball covers forward thinking, theory of mind, behavioral flexibility, the ability to make plans, and complex vocal communications in the animal kingdom. Readers also learn about the hive minds of termites and bees.



The Book of Minds
Philip Ball
University of Chicago Press, 2022. 512 pp.

Ball draws special attention to cephalopods, whose minds are as different from ours as it gets. Prospecting minds further in the living world, he makes a brief foray into "plant neurobiology" and even entertains James Lovelock's Gaia hypothesis, which posits that our planet is a self-regulating living organism.

Ball also tackles artificial minds, as well as our projections upon them. Does the claim that machines do not or cannot have minds reenact the bias against animals that has taken us so long to correct, he wonders? While it is true that computers do specific tasks better and faster than us, the recurring promise that sentient artificial intelligence is just one line of code away has grown tiresome. Good-enough mimicry does not a mind make.

It is unlikely that all the minds in the Universe are confined to this pale blue dot. How will we know aliens when we meet them? Ball suggests that our encounter with extraterrestrials is more likely to be technological than biological, that is, through their gadgets rather than themselves.

And what about the mind of God? It certainly defies mapping. Ball does not believe in such a thing, and yet—in contrast to the easy ridiculing of certain vociferous atheists—he respects the subtle work of theologians.

A paradox awaits the reader toward the book's end: Did they make it there by choice? In chapter 9, Ball dares to tackle free will. If it exists (he is inclined to think it does), "minds alter the universe in an astonishing way," he writes. If it does not, do minds matter? Quantum indeterminacy and classical determinism offer chance and necessity, while the words "free" and "will" suggest a mysterious force that defies physical possibility. And yet, the minds that laws permit must ultimately face the laws that minds admit.

In considering all possible minds, *The Book of Minds* suggests that humankind is not at the center, at least spatially. However, the book risks framing other minds as surrogates to explore our own. Pluralism and decentering are desirable, but human imagination bootstraps us above other creatures.

Given the scope and length of the book, it is surprising that the phenomenology of experience, the consciousness-expanding effects of psychedelic substances, and altered states of mind such as lucid dreaming and near-death experiences are only mentioned in passing. On these and other fringe topics, Ball reaches the boundaries of the scientific orthodoxy but never crosses the line.

Ball ends the book with a final plea for why we should concern ourselves with the concept of minds, insisting that such probing "might show us what we can become." The French priest and paleontologist Pierre Teilhard de Chardin would have likely agreed. Humanity has not only evolved, he once argued, it is evolving. It is not really us who know the Universe; the Universe knows itself through us. ■

10.1126/science.abq6975

The reviewer is at the Instituto de Neurociencias, Consejo Superior de Investigaciones Científicas–Universidad Miguel Hernández de Elche, Alicante, Spain. Email: agomezmarin@gmail.com

PALEOANTHROPOLOGY

Discovering the "Hobbit"

A paleoanthropologist unpacks the curious tale of the diminutive hominin *Homo floresiensis*

By **Richard G. Roberts** and **Thomas Sutikna**

On 28 October 2004, *Homo floresiensis* became a scientific and media sensation. Dubbed the "Hobbit" after J. R. R. Tolkien's wee folk, this primitive little hominin was thought to have survived on the island of Flores in Indonesia until just 18,000 years ago (subsequently revised to 50,000 years ago). *H. floresiensis* stood only 1 meter tall and had several odd features, including a small head, hunched shoulders, short legs, and long flat feet with curved toes. The news was greeted with glee, astonishment, skepticism, and counterclaims that the remains were of a diseased member of our own species. In *Little Species, Big Mystery*, the story of *H. floresiensis* is brought to life by Australian paleoanthropologist Debbie Argue.

Argue's book recounts the events leading up to the announcement of the discovery of the Hobbit, the controversy in its aftermath, and the subsequent research into the archaeology, chronology, and ancestry of this peculiar hominin. Recent discoveries of ancient remains of other diminutive humans on Flores and in the Philippines, and suggestions that Denisovans—another group of extinct hominins—may have ventured into island Southeast Asia, have shone a spotlight on this region as an intriguing field experiment in human evolution.

Ancient DNA has yet to be recovered from hominin remains older than a few thousand years in the tropics, and the use of ancient proteins to elucidate hominin taxonomy is in its infancy. Revealing the history of the Hobbit, its contemporaries, and their ancestors therefore depends on detailed analysis of the few and typically fragmentary bones and teeth preserved over the eons. Argue takes care to demystify anatomical terms for the uninitiated and adopts an open-minded but sensibly cautious approach to interpreting the meager human fossil record, following in the footsteps of her colleague and mentor, mammal taxonomist Colin Groves. Together with others, they were the first to independently confirm *H. floresiensis* as a valid new species.

Argue debunks the various ailments attributed to the Hobbit, which its detractors considered a *H. sapiens* with microcephaly, Laron syndrome, cretinism, or Down syndrome. Only the first of these was in play when Mike Morwood, coleader of the Indonesian–Australian research team, and Penny van Oosterzee released *The Discovery of the Hobbit* in 2007, so Argue's book provides a welcome update, benefiting from her role as a researcher who was not embroiled in the initial controversy and a perspective that only the passage of time and additional data can bring.

Argue also grapples with questions that are currently unresolved, such as the position of the Hobbit in the hominin family tree and the routes and means by which its ancestors reached Flores. The island has always been separated from mainland Asia, so these small-brained hominins must have made several sea crossings to achieve landfall and then persisted for several tens or hundreds of millennia until their final disappearance.

Argue contemplates various means of travel but defers judgment, given the lack of available evidence either way. She also considers the jury to still be out on whether these ancient mariners arrived from islands to the north or west of Flores. She is satisfied, however, that the available skeletal evidence supports the evolution of Hobbits from small-bodied hominins who dispersed out of Africa more than 2 million years ago, rather than from a large-bodied ancestor (*H. erectus*) who dwarfed over time in response to environmental pressures on Flores, which is the alternative hypothesis still in contention.

The absence of fossils of little, primitive humans between Africa and Flores might appear to weaken Argue's case, but hominin fossils in South Asia—even those of *H. sapiens*—are exceedingly scarce. However, two recent finds may hold clues to this outstanding question. East of the home of the Hobbit on Flores, a few fragmentary remains of a 700,000-year-old tiny hominin were unearthed in 2014. Were these the ancestors of the Hobbit? Too few fossils to tell, Argue opines. Meanwhile, in the Philippines, the remains of a small, late-surviving hominin—*H. luzonensis*—suggest an intriguing parallel with the Hobbit. How closely were the two diminutive species related? More data from more sites are needed, she writes.

While she is optimistic that ancient DNA can help elucidate the evolutionary relationships of *H. floresiensis* and other hominins in island Southeast Asia, what we ultimately need most are more human fossils, maintains Argue in a clarion call for local researchers, who remain central to the Flores and Philippines projects, and their international collaborators. ■

10.1126/science.abq8248

The reviewers are at the Centre for Archaeological Science and the Australian Research Council Centre of Excellence for Australian Biodiversity and Heritage, University of Wollongong, Wollongong, NSW 2522, Australia. Email: rgrob@uow.edu.au; thomasutikna@gmail.com

Little Species, Big Mystery: The Story of *Homo Floresiensis*
Debbie Argue
Melbourne University Press, 2022. 216 pp.

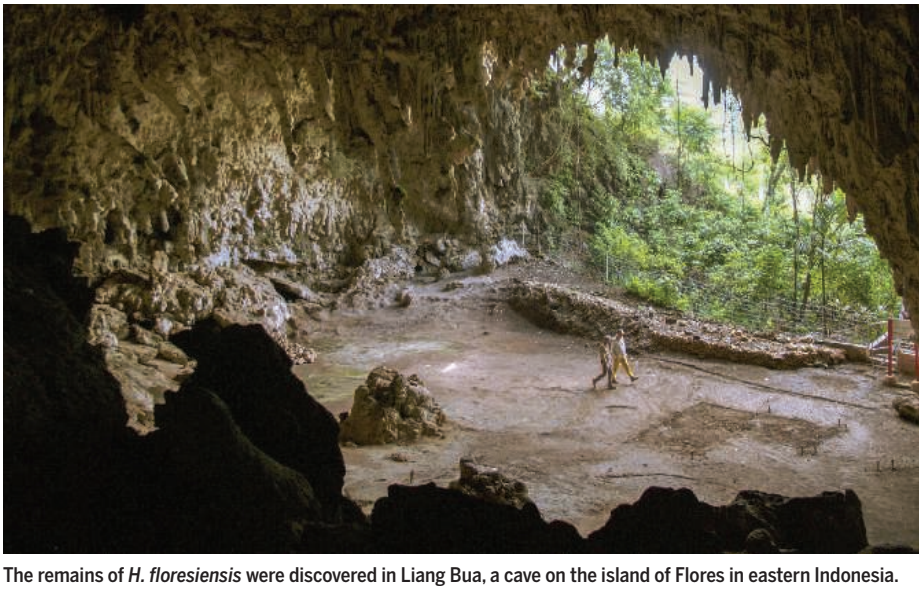


PHOTO: JAVIER TRUEBA/MSF/SCIENCE SOURCE

The remains of *H. floresiensis* were discovered in Liang Bua, a cave on the island of Flores in eastern Indonesia.

The Cerrado region, including this spring in the state of Minas Gerais, is crucial to Brazil's water security.

Edited by **Jennifer Sills**

Cerrado conservation is key to the water crisis

In 2021, the hydroelectric plant reservoirs in southern and midwestern Brazil, which provide 70% of the country's hydraulic energy (*1*), reached their lowest levels in 91 years (*1*). In addition to compromising the country's electricity sources, declining water levels threaten Brazil's population by reducing the available food and water supply of food, which drives up prices and affects both family budgets and trade (*2*). Drought also exacerbates forest fires, disrupts the life cycles and distribution of local species, and limits the productivity of ecosystems. To address these challenges, Brazil must work to protect its water supply by conserving the Cerrado region.

Covering more than 2 million square kilometers, the Cerrado supplies water to 8 of Brazil's 12 hydrographic regions (*3*). Despite its importance, only about 8% of the territory is legally protected (*4*). Agribusinesses have replaced the Cerrado's native vegetation with extensive areas of monoculture and livestock (*5*), which use about 70% of the region's water (*6*). Deforestation in the region has reduced the region's rainfall, affecting the local climate and hindering the recharge of aquifers (*5*, *7*, *8*).

Despite warnings from the scientific community, Brazil continues its relentless pace toward a harsh future in which water and food security will be under threat. In addition to saving the Amazon rainforest, Brazil must urgently take steps to protect the Cerrado. Given that protected areas in the region are an effective conservation device (*9*), more should be designated. Conservation measures should focus on the Cerrado's water sources, including headwaters of rivers. Brazil should limit deforestation and invest instead in sustainable agriculture, using only existing agricultural fields. Finally, the country should amend the Brazilian Forest Code to increase the protected area within rural properties located in the Cerrado (*10*). The loss of the Cerrado region's biodiversity and water production could be catastrophic.

Cássio Cardoso Pereira* and Geraldo Wilson Fernandes
Programa de Pós-Graduação em Ecologia, Conservação e Manejo da Vida Silvestre, Departamento de Genética, Ecologia, e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, 31270-910, Minas Gerais, Brazil.
*Corresponding author.
Email: cassiocardosopereira@gmail.com

REFERENCES AND NOTES

1. W. Brandimarte, G. Freitas Jr., "Brazil's worst water crisis in 91 years threatens power supplies," *Bloomberg* (2021).
2. "Water crisis in Brazil: What are the impacts on the economy?," *ParMais* (2021); www.parmais.com.br/blog/crise-hidrica-no-brasil/ [in Portuguese].
3. M. Lahsen, M. M. Bustamante, E. L. Dalla-Nora, *Environ. Sci. Pol. Sustain. Dev.* **58**, 4 (2016).
4. L. L. Rausch *et al.*, *Conserv. Lett.* **12**, e12671 (2019).
5. G. W. Fernandes *et al.*, *Perspect. Ecol. Conserv.* **14**, 146 (2016).
6. G. W. Fernandes *et al.*, *Cerrado: Mm Busca de Soluções Sustentáveis* (Editora Vozes, Rio de Janeiro, 2016).
7. J. A. Rodrigues *et al.*, *Int. J. Climatol.* **40**, 2511 (2020).
8. A. L. D. Silva *et al.*, *Water* **13**, 1054 (2021).
9. T. Carranza, A. Balmford, V. Kapos, A. Manica, *Conserv. Lett.* **7**, 216 (2014).
10. J. P. Metzger *et al.*, *Perspect. Ecol. Conserv.* **17**, 91 (2019).

10.1126/science.add4719

Zero-COVID Editorial lacks balance

It is one thing to hope for an end to China's zero-COVID policy. It is quite another to critique it without sensitivity to the issues. In his Editorial "Zeroing out on zero-COVID" (3 June, p. 1026), W. C. Kirby ostensibly pleads for better conditions to reopen scholarly exchange but seems more interested in accusing China of supposed failures.

Kirby blames China for not doing enough to contain the disease initially, only to accuse the country of now being too rigorous. When considering how China might transition to a different policy, he does not appreciate the difficulties of doing so in Australia, Canada, New Zealand, and Singapore [e.g., (*1*–*3*)], nor does he acknowledge the challenges unique to China, such as disparities in urban and rural areas, lack of education, low vaccination rates among the elderly, and a culture tolerant of short-term sacrifices for long-term policy gains (*4*).

Kirby claims that because the Chinese approach presents difficulties in hindsight, it was wrong to adopt it in the first place; he fails to acknowledge that a decision can be made in good faith that later has adverse effects. Kirby also alleges that this supposed error in judgment was the result of politics interfering with scientists, who would never be foolish enough to recommend a zero-COVID approach. In fact, this approach has frequently been discussed as a credible alternative (*2*). By making these assumptions, Kirby undermines rather

PHOTO: CÁSSIO CARDOSO PEREIRA

than promotes academic dialogue with our colleagues in China.

Carl Mitcham[1]*, Alfred Nordmann[2], Yongmou Liu[3]
[1]Department of Liberal Arts and International Studies, Colorado School of Mines, Golden, CO 80401, USA. [2]Institut für Philosophie, Technische Universität Darmstadt, Darmstadt, Hessen, Germany. [3]Renmin University of China, Beijing, China.
*Corresponding author.
Email: cmitcham@mines.edu

REFERENCES AND NOTES

1. "Asian countries are at last abandoning zero-covid strategies," *The Economist* (2021).
2. A. Llupià *et al*., "What is a zero-COVID strategy and how can it help us minimise the impact of the pandemic?" (Barcelona Institute for Global Health, 2020).
3. J. Curtin, "The end of New Zealand's zero-COVID policy," *Think Global Health* (2021).
4. Z. Jun, "What justifies China's zero-COVID policy?," *Project Syndicate* (2022); www.project-syndicate.org/commentary/shanghai-lockdown-why-china-keeps-its-zero-covid-strategy-by-zhang-jun-2022-05.

10.1126/science.add5130

Vaccinate to combat COVID-19 in China

As W. C. Kirby argues in his Editorial "Zeroing out on zero-COVID" (3 June, p. 1026), the cost of China's strict zero-COVID policy is enormous. From local to regional to international impacts, this policy is likely to increase unemployment and poverty while further disrupting global supply, communication, and cooperation. China should replace its zero-COVID strategy with a plan to increase vaccination rates.

This year marks the largest surge in COVID-19 cases in China since the beginning of the outbreak in 2019 (*1*). Shanghai was the epicenter of the surge, recording 625,186 positive cases during the city's strict lockdown from 1 April to 1 June (*2*). Most cases (90.75%) were asymptomatic, and the death rate was only 0.094% (*2*). Of the 588 people in Shanghai who died from Covid-19 during this time, the average age was 82.9, and all had severe preexisting medical conditions (*3*). For those above 80, the vaccination rate was a meager 1.3% (*3*).

Between 10 March and 16 April, each Shanghai resident underwent more than a dozen polymerase chain reaction (PCR) tests, in addition to daily at-home antigen testing (*4*). Long lines for testing increased the possibility of spreading the virus, and with tens of millions of people taking tests daily, false positives were frequent (*5*). The cost of massive testing sites and test materials, combined with the required employee salaries, could reach billions of US dollars (*6*).

Lockdowns and mandatory proactive testing have high social and economic costs as well. People have had difficulty accessing hospital services, acquiring medications, and making routine medical appointments (*7*). Loss of employment has led to food and housing insecurity (*8*). Shanghai's industrial output fell by 61.5% in April compared to a year earlier. Retail sales fell by 48.3% over the same period (*9*).

Instead of spending billions on testing and quarantine, China should invest in vaccination (*10*). Data from Shanghai show that with three doses of China's vaccine, including two primary shots and a booster, the incidence of severe symptoms decreases by 90% (*11*). Yet vaccination rates remain surprisingly low. By May, only 62% of Shanghai residents older than 60 had received two doses of the vaccine, and only 39% had received boosters. The cost of vaccinating all residents over 60 would be much lower than the cost of regular testing (*12*), while mitigating the social and economic consequences of the zero-COVID strategy.

Ruishan Chen[1]*, Xiaona Guo[1], Annah Zhu[2]
[1]Shanghai Jiaotong University, Shanghai 200040, China. [2]Wageningen University, 6706KN Wageningen, Netherlands.
*Corresponding author.
Email: chenrsh04@gmail.com

REFERENCES AND NOTES

1. "Shanghai starts China's biggest COVID-19 lockdown in 2 years," *Associated Press* (2022).
2. Shanghai Municipal Health Commission, "Shanghai COVID-19 cases update" (2022); https://wsjkw.sh.gov.cn/yqtb/index.html [in Chinese].
3. "Severe symptoms caused by Omicron, how to save the patients?," CCTV (2022); https://tv.cctv.cn/2022/05/13/VIDEpNpQ943Uq9zoDq3RuI1Q220513.shtml [in Chinese].
4. "Shanghai completed 13 polymerase chain reaction (PCR) tests from March 10th to April 16th, testing more than 200 million people total," CCTV (2022); https://cn.chinadaily.com.cn/a/202204/17/WS625c02baa3101c3ee7ad0db6.html [in Chinese].
5. L. Jialin, "COVID-19 polymerase chain reaction PCR test false positive events," *The Economic Observer* (2022); www.eeo.com.cn/2022/0514/534378.shtml [in Chinese].
6. A. Stevenson, "For its next zero covid chapter, China turns to mass testing," *The New York Times* (2022).
7. M. Guanjun, "During the pandemic, the road to see a doctor for hemodialysis patients in Shanghai," *Shanghai Observer* (2022); www.jfdaily.com/news/detail?id=480318 [in Chinese].
8. V. Wang, "'I'm very anxious': China's lockdowns leave millions out of work," *The New York Times* (2022).
9. E. Zhang, R. Woo, "Shanghai economy hit on all sides in April by COVID lockdown," *Reuters* (2022).
10. X. Zhang *et al*., *Lancet* **399**, 2011 (2020).
11. Shanghai Municipal Health Commission, "COVID situation press conference of Shanghai" (2022); http://wsjkw.sh.gov.cn/xwfb/20220528/82acef0965d245609311e5448991f8f2.html [in Chinese].
12. "Many cities turn to mass testing, how to do that effectively," *Xinhua News* (2022); www.news.cn/politics/2022-06/08/c_1128723529.htm [in Chinese].

10.1126/science.add4602

RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk

COSMOCHEMISTRY

Mars accreted in an unexpected order

The Solar System formed from a nebula of gas and dust, from which the Sun and planets accreted. Their assembly sequence can potentially be reconstructed using the abundances of chemical elements and their isotope ratios. Péron and Mukhopadhyay have measured krypton and xenon isotopes in the Martian meteorite Chassigny, already known to reflect the planet's interior composition. They found isotope ratios similar to meteorites, indicating that much of Mars' mass was sourced from solids. However Mars' atmosphere has isotopes indicating an origin from the nebular gas and thus must have been added later than the interior. Standard cosmochemical models predict the opposite sequence, so the results challenge current understanding of Solar System assembly. —KTS

Science, abk1175, this issue p. 320

Satellite image reconstruction of the surface of Mars

QUANTUM SIMULATION

Simulating thermalization dynamics

Calculating the dynamics of gauge theories, which underlie some of the most successful models in physics, is extremely challenging for classical computers. An alternative to computation is quantum simulation using tunable physical systems in which gauge symmetry constraints can be effectively implemented. Zhou *et al*. studied the thermalization of a U(1)-symmetric gauge theory using cold bosonic atoms trapped in a tilted staggered optical lattice. The system's evolution depended on whether the gauge constraint was enforced. Additionally, different gauge-symmetric initial states with the same energy density evolved to the same thermal state. —JS

Science, abl6277, this issue p. 311

CELL BIOLOGY

Predicting differentiation with morphology

Obtaining a differentiated, functionally specialized state is accompanied by alterations in cell morphology. Huang *et al.* report that analysis of morphology-related gene expression can be used to define stem cell differentiation stages in a model of in vitro surface epithelium differentiation from pluripotent stem cells. Analysis of morphology-related gene expression, in combination with chromatin conformation and epigenetic state, further identified a novel transcriptional master regulator, GRHL3, that drives the initiation of surface epithelium lineage commitment in collaboration with the growth factor BMP4 by increasing chromatin accessibility. These findings underline the tight relationships between cell morphology and state and further define collaborative actions of transcription factors and morphogens. —SAW

Sci. Adv. 10.1126/sciadv.abo5668 (2022).

WATER STRUCTURE

How hydrated protons gather at interfaces

Hydrated protons play a critical role in numerous physical and chemical processes. However, their atomic-level characterization remains experimentally challenging, especially for interfacial systems. Using qPlus-based atomic force microscopy and path integral molecular dynamics, Tian *et al*. demonstrated real-space imaging of the most representative forms of hydrated protons, the Zundel and the Eigen cations, at water networks grown on Au(111) and Pt(111) surfaces. The authors characterized the preferable two dimensional assemblies of these cations at the interfaces (see the Perspective by Sugimoto). They also revealed an interesting mechanism of the coupled Eigen-Zundel interconversion involving interfacial proton transfer, which may well be an important pathway in

PHOTO: NASA/JPL-CALTECH/USGS

electrochemical processes such as hydrogen evolution. —YS

Science, abo0823, this issue p. 315; see also add0841, p. 264

IMMUNOLOGY

Sunburn: A ZAK attack

Nod-like receptors (NLRs) are intracellular sensors of infection that, when activated, assemble into signaling complexes (inflammasomes) that induce proinflammatory cytokine production and pyroptotic cell death. One such NLR, NLRP1, is expressed in the skin and can be triggered by ultraviolet B (UVB) radiation, the cause of acute sunburn. Robinson *et al*. report that UVB-induced RNA photolesions stall ribosomes, activating the ribotoxic stress response kinase ZAKα. Together with p38, ZAKα directly phosphorylates the disordered linker region of human NLRP1. Several microbial ribotoxins also trigger this pathway, suggesting that it may serve as a target for future anti-inflammatory therapies. —STS

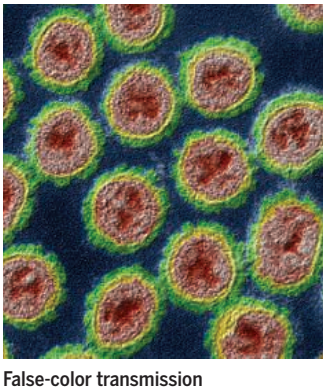
Science, abl6324, this issue p. 328

INFLUENZA

A versatile vaccine

A "universal" vaccine for influenza A virus (IAV) is a high priority to protect against seasonal epidemics and for pandemic preparedness. Park *et al.* developed a multivalent, inactivated whole-virus vaccine for IAV. The vaccine included four IAV subtypes that conferred protection against multiple antigenically distinct and fully heterosubtypic IAV strains in mice and ferrets. These data support further clinical development of this vaccine. —CSM

Sci. Transl. Med. **14**, eabo2167 (2022).



False-color transmission electron microscope image of avian influenza virus

PHOTOS (LEFT TO RIGHT): EYE OF SCIENCE/SCIENCE SOURCE; DAMOCEAN/ISTOCK.COM

MOLECULAR SEPARATION

Using polymers to separate xylene isomers

Each of the three xylene isomers is a valuable feedstock for many chemical processes. However, because these isomers have similar boiling points, separation by distillation or other evaporative methods is energy intensive and inefficient. Li *et al*. developed an alternative separation process using a manganese-based, stacked, one-dimensional coordination polymer with an interchain distance that varies with the temperature and content of the adsorbed hydrocarbons. At 393 kelvin, only *p*-xylene can access the voids, whereas *m*-xylene can enter at temperatures between 333 and 393 kelvin and *o*-xylene only below 333 kelvin. Through a sequence of separations, each isomer can be extracted with high selectivity. The material is simple, cost-effective, and easily scalable, and it exhibits remarkable water and air stability and excellent recyclability. —MSL

Science, abj7659, this issue p. 335

SOLAR CELLS

A fluoride boost

The wide-bandgap perovskite layer in perovskite-silicon tandem solar cells is still limited by high interface recombination at the electron extraction interface. Liu *et al*. show that adding an ultrathin magnesium fluoride interlayer between the perovskite and C_{60} electron transport layer during growth facilitates mitigated nonradiative recombination. An analysis of electronic structural data showed that conduction band bending of the perovskite and C_{60} facilitated electron extraction. A monolithic perovskite-silicon tandem solar cell with a certified power conversion efficiency of 29.3% retained about 95% of its initial performance for 1000 hours. —PDS

Science, abn8910, this issue p. 302

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

VISUAL SYSTEMS

Seeing things differently

We see the world through our own eyes and generally presume that what we see is "reality." In fact, human visual systems see only the parts of the world that they have evolved to see, and different species see the world differently. Murphy and Westerman looked across published data for patterns of wavelength sensitivity across invertebrates and vertebrates. It seems that an animal's environment shapes its mode of seeing, such that species that have evolved in different terrestrial or aquatic habitats, or even closed versus open canopy habitats, have different light wavelength sensitivities. However, the authors also found that phylum appears to limit the degree to which sensitivity can be shaped, indicating that evolutionary history may be hard to overcome. —SNV *Proc. Biol. Sci.* **289**, 20220612 (2022).

Species living in different habitats, such as fish and birds, have evolved visual systems with different light wavelength sensitivities.

PROTEIN AGGREGATION

A chemical chaperone in the lens

The eye's lens is almost entirely composed of specialized proteins called crystallins that must persist intact over the course of a person's life. Aggregation results in cataracts and degraded vision. Serebryany *et al.* used a combination of biochemical and biophysical techniques to show that the abundant lens metabolite *myo*-inositol suppresses aggregation

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

Importance of infection history

A long-term study of health care workers in the United Kingdom has allowed their history of infection and vaccination to be traced precisely. Reynolds *et al*. found some unexpected immune-damping effects caused by infection with a heterologous variant to the latest wave of infection by the Omicron/Pango lineage B.1.1.529. The authors found that Omicron infection boosted immune responses to all other variants, but responses to Omicron itself were muted. Infection with the Alpha variant provided weaker boosting for Omicron-specific responses. Furthermore, Omicron infection after previous Wuhan Hu-1 infection failed to boost neutralizing antibody and T cell responses against Omicron, revealing a profound imprinting effect and explaining why frequent reinfections occur. —CA

Science, abq1841, this issue p. 275

CANCER IMMUNOLOGY

TCR usage defines γδ T cell response

Gamma-delta (γδ) T cells are a subtype of T lymphocytes that perform epithelial surveillance and orchestrate pro- and anti-inflammatory responses. They have been shown to have divergent roles in tumor development, with the general view being that the interleukin-17 (IL-17)–producing subset can promote tumor growth, whereas cytotoxic interferon-γ (IFN-γ)–producing cells can protect against tumors. Reis *et al.* now add additional layers of understanding to how the opposing functions of IL-17$^+$ versus IFN-γ$^+$ intestinal γδ subsets affect colon cancer. Using mouse models, tumor-infiltrating γδ T cells were found to assist tumor growth, whereas tissue-resident cells could fight tumors. Defined T cell receptor Vγδ gene usage appears key to the functional differences of γδ T cells: Infiltrating pro-tumorigenic cells use Vγ4/Vγ6, whereas resident anti-tumorigenic cells use Vγ1/Vγ7. —PNK

Science, abj8695, this issue p. 276

ASTEROIDS

Grabbing a sample of asteroid Bennu

The near-Earth carbonaceous asteroid Bennu was the target of the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, Security–Regolith Explorer) sample return mission. After rendezvousing with the asteroid, the spacecraft spent 2 years studying its surface and selecting a suitable site. Lauretta *et al*. describe the sample collection process and its effects on Bennu. The asteroid provided almost no resistance to contact, and the gas released by the spacecraft blew a crater several meters wide, exposing redder rocks and dust. So much material was gathered that the collection chamber overflowed. Approximately 250 grams was successfully stowed, well above the mission goal of 60 grams. The samples are expected to arrive on Earth in September 2023. —KTS

Science, abm1018, this issue p. 285

HEART DISEASE

A disheartening loss

Although the Y chromosome is the smallest and contains few genes, its functions are not fully understood. It has been observed, however, that mosaic loss of the Y chromosome in blood cells frequently occurs with age, and this alteration is associated with various medical conditions. Sano *et al*. modeled this process in mice by reconstituting their bone marrow with cells lacking the Y chromosome (see the Perspective by Zeiher and Braun). The resulting mice were prone to fibrosis and decreased cardiac function, especially in the setting of pressure overload, but they benefited from treatment with a transforming growth factor β1–neutralizing antibody. Human patients with loss of chromosome Y in their blood were also at greater risk of cardiac pathology, supporting the clinical relevance of these findings. —YN

Science, abn3100, this issue p. 292; see also add0839, p. 266

SURFACE CHEMISTRY

Tip-induced organic reactions

Control over the reaction products of a unimolecular transformation on a surface have been induced and visualized with a scanning tunneling microscope (STM) tip. Albrecht *et al*. synthesized a tetrachlorotetracene molecule and absorbed it on a thin salt layer grown on copper (see the Perspective by Alabugin and Hu). Under cryogenic conditions, voltage pulses from the STM tip led to the elimination of the chlorine atoms and produced intermediates with a large central ring. Subsequent voltage pulses created other isomers of this molecule, a diyne and a chrysene-based bisaryne, in reactions that could be reversed with opposite polarity pulses. —PDS

Science, abo6471, this issue p. 298; see also abq2622, p. 261

SOLAR CELLS

A cap against aging

Accelerated aging tests for perovskite solar cells must take into account several degradation pathways. Zhao *et al*. found that for all-inorganic cesium lead triiodide ($CsPbI_3$) solar cells, a two-dimensional $Cs_2PbI_2Cl_2$ capping layer stabilized the interface between the $CsPbI_3$ absorber and the copper thiocyanate hole-transporter layer and boosted its power conversion efficiency to 17.4% (see the Perspective by Habisreutinger and Reese). Accelerated testing at various temperatures up to 110°C and approximately 65% relative humidity revealed an Arrhenius temperature dependence on efficiency degradation. The solar cell should maintain 80% of its efficiency for more than 5 years at 35°C. —PDS

Science, abn5679, this issue p. 307; see also abq7013, p. 265

MEDICINE

Drugs in the environment

Pharmaceutical pollution, when drugs used to treat humans and animals contaminate the environment, can disrupt ecosystems and lead to diverse effects in various organisms. The increased use of drugs, especially during epidemics and pandemics, allows more pharmaceutical products to enter the environment, usually through wastewater. In a Perspective, Orive *et al.* discuss the problem of pharmaceutical pollution and its impacts on ecosystems. They identify key areas that need further research and development, including decontaminating wastewater, understanding the impacts of drug contaminants, developing "greener drugs" that are more easily eliminated, and educating health-care practitioners about the impact of drug pollution. —GKA

Science, abp9554, this issue p. 259

NEUROSCIENCE

Fine-tuning information transfer

To generate adaptive behavior, our brains constantly combine information from multiple sources. How do neuronal circuits orchestrate and maintain the balance of different input streams in the face of constant change? Sakalar *et al*. discovered that neurogliaform cells were strongly coupled with gamma oscillations that are associated with gating the

interaction of hippocampus and cortex (see the Perspective by Craig and Witton). The activity of neurogliaform cells was correlated with a decrease in coupling between pyramidal cell firing and gamma oscillations without affecting the overall levels of activity of the pyramidal cells. Neurogliaform cells locally released the neurotransmitter γ-aminobutyric acid, which selectively decreased the influence of neocortical inputs to hippocampal area CA1 at specific stages in the local field potential. This modulation of inputs allows for the transfer of different types of information at different times. —PRS

Science, abo3355, this issue p. 324; see also add2681, p. 262

DEVELOPMENTAL BIOLOGY

Amplifying gamete development

Mutations in the X chromosome gene *RLIM*, which encodes the ubiquitin ligase RNF12, cause hypogenitalism in male patients and infertility in female carriers. Using *Rlim*$^{-/y}$ mouse embryonic stem cells reconstituted with human RNF12, Segarra-Fas *et al*. identified a transcriptional feed-forward loop by which RNF12 activity amplified the derepression of genes required for gametogenesis, thereby driving germ cell differentiation. *RLIM* and *USP26* mutations associated with infertility in patients disrupted this system. —AMV

Sci. Signal. **15**, eabm5995 (2022).

MUCOSAL IMMUNOLOGY

Custom-tailored gut IgA

Much of the secretory immunoglobulin A (IgA) antibody present in the gut lumen is reactive with bacterial antigens, but the degree of specificity that this IgA displays across the diverse gut microbiome has not been established. Using germ-free mice recolonized with single bacterial strains or defined consortia of commensal bacteria, Yang *et al*. investigated the degree of cross-reactivity exhibited by polyclonal fecal and serum IgA or by monoclonal IgA antibodies from IgA hybridoma clones derived from lamina propria B cells. Most IgAs exhibited specificity for the inducing species of gut bacteria, with only a minor fraction of monoclonal IgAs showing broader reactivity. These findings suggest that the bulk secretory IgA repertoire in the gut is personalized to match the community of commensal bacteria residing in the host's intestine. —IRW

Sci. Immunol. **7**, eabg3208 (2022).

electrochemical processes such as hydrogen evolution. —YS

Science, abo0823, this issue p. 315; see also add0841, p. 264

IMMUNOLOGY

Sunburn: A ZAK attack

Nod-like receptors (NLRs) are intracellular sensors of infection that, when activated, assemble into signaling complexes (inflammasomes) that induce proinflammatory cytokine production and pyroptotic cell death. One such NLR, NLRP1, is expressed in the skin and can be triggered by ultraviolet B (UVB) radiation, the cause of acute sunburn. Robinson *et al*. report that UVB-induced RNA photolesions stall ribosomes, activating the ribotoxic stress response kinase ZAKα. Together with p38, ZAKα directly phosphorylates the disordered linker region of human NLRP1. Several microbial ribotoxins also trigger this pathway, suggesting that it may serve as a target for future anti-inflammatory therapies. —STS

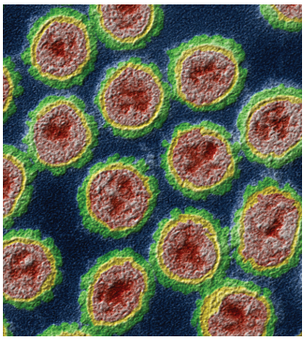
Science, abl6324, this issue p. 328

INFLUENZA

A versatile vaccine

A "universal" vaccine for influenza A virus (IAV) is a high priority to protect against seasonal epidemics and for pandemic preparedness. Park *et al.* developed a multivalent, inactivated whole-virus vaccine for IAV. The vaccine included four IAV subtypes that conferred protection against multiple antigenically distinct and fully heterosubtypic IAV strains in mice and ferrets. These data support further clinical development of this vaccine. —CSM

Sci. Transl. Med. **14**, eabo2167 (2022).



PHOTOS (LEFT TO RIGHT): EYE OF SCIENCE/SCIENCE SOURCE; DAMOCEAN/ISTOCK.COM

False-color transmission electron microscope image of avian influenza virus

MOLECULAR SEPARATION

Using polymers to separate xylene isomers

Each of the three xylene isomers is a valuable feedstock for many chemical processes. However, because these isomers have similar boiling points, separation by distillation or other evaporative methods is energy intensive and inefficient. Li *et al*. developed an alternative separation process using a manganese-based, stacked, one-dimensional coordination polymer with an interchain distance that varies with the temperature and content of the adsorbed hydrocarbons. At 393 kelvin, only *p*-xylene can access the voids, whereas *m*-xylene can enter at temperatures between 333 and 393 kelvin and *o*-xylene only below 333 kelvin. Through a sequence of separations, each isomer can be extracted with high selectivity. The material is simple, cost-effective, and easily scalable, and it exhibits remarkable water and air stability and excellent recyclability. —MSL

Science, abj7659, this issue p. 335

SOLAR CELLS

A fluoride boost

The wide-bandgap perovskite layer in perovskite-silicon tandem solar cells is still limited by high interface recombination at the electron extraction interface. Liu *et al*. show that adding an ultrathin magnesium fluoride interlayer between the perovskite and C_{60} electron transport layer during growth facilitates mitigated nonradiative recombination. An analysis of electronic structural data showed that conduction band bending of the perovskite and C_{60} facilitated electron extraction. A monolithic perovskite-silicon tandem solar cell with a certified power conversion efficiency of 29.3% retained about 95% of its initial performance for 1000 hours. —PDS

Science, abn8910, this issue p. 302

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

VISUAL SYSTEMS

Seeing things differently

We see the world through our own eyes and generally presume that what we see is "reality." In fact, human visual systems see only the parts of the world that they have evolved to see, and different species see the world differently. Murphy and Westerman looked across published data for patterns of wavelength sensitivity across invertebrates and vertebrates. It seems that an animal's environment shapes its mode of seeing, such that species that have evolved in different terrestrial or aquatic habitats, or even closed versus open canopy habitats, have different light wavelength sensitivities. However, the authors also found that phylum appears to limit the degree to which sensitivity can be shaped, indicating that evolutionary history may be hard to overcome. —SNV *Proc. Biol. Sci.* **289**, 20220612 (2022).

Species living in different habitats, such as fish and birds, have evolved visual systems with different light wavelength sensitivities.

PROTEIN AGGREGATION

A chemical chaperone in the lens

The eye's lens is almost entirely composed of specialized proteins called crystallins that must persist intact over the course of a person's life. Aggregation results in cataracts and degraded vision. Serebryany *et al*. used a combination of biochemical and biophysical techniques to show that the abundant lens metabolite *myo*-inositol suppresses aggregation

TROPICAL ECOLOGY

Fire shapes a mosaic landscape

Tropical savannas, grassy ecosystems with sparse trees, often occur in a mosaic of patches interspersed with forest. In Southeast Asia, most savannas are misclassified as forests, and the factors that determine where they occur are poorly understood. Pletcher *et al.* used remotely sensed data to identify drivers of tree density and the distribution of forest-savanna mosaics across Southeast Asia. They found that lower tree cover (savannas) occurred in areas with less and more seasonal rainfall and higher fire frequency, but drier landscapes with intermediate fire frequency had more discrete patches of savanna versus forest. The importance of fire for maintaining savannas indicates that current practices of fire suppression could negatively affect this ecosystem. —BEL

Ecography 10.1111/ecog.06280 (2022).

A combination of rainfall and fire creates patterns of wooded landscapes in Southeast Asia.

of otherwise susceptible lens crystallin variants. This effect was also seen for some other common sugars and sugar alcohols. Experiments suggest that inositol functions by inhibiting the early, bimolecular kinetic step in aggregation rather than by stabilizing the native state or destabilizing the aggregate. —MAF

eLife **11**, e76923 (2022).

STELLAR PHYSICS

A magnetometer for stellar cores

Asteroseismology, the study of stars' vibrational frequencies, provides information on stellar interiors. High magnetic fields in a star's core can suppress the appearance of low-frequency modes. Determining which frequencies are missing can therefore be used to measure core magnetic fields, an observational technique that had been applied only to red giant stars. Lecoanet *et al.* have extended this method to HD 43317, a nearby high-mass main sequence star. They inferred a core magnetic field of about 450 kilogauss, higher than expected given the star's previously measured surface field of 1.3 kilogauss. The results constrain the origin of magnetic fields in highly magnetized main sequence stars. —KTS

Mon. Not. R. Astron. Soc. **512**, L16 (2022).

CELL BIOLOGY

Bypassing clathrin using curvature

Clathrin-mediated endocytosis involves the generation of membrane invaginations that pinch off to internalize vesicles from the cell surface. Clathrin provides a mechanical coat that induces and stabilizes membrane curvature. Cail *et al*. wanted to know if the role of clathrin could be bypassed if the plasma membrane were forced into the correct shape. They used nanopatterning to produce glass-like substrates with U-shaped features mimicking the membrane invaginations observed during clathrin-mediated endocytosis. The plasma membranes of cells grown on these surfaces curved to conform to the invaginations, which were able to recruit the endocytic machinery downstream of clathrin. This allowed for endocytic site maturation and partially bypassed clathrin's role in endocytosis. The authors conclude that clathrin's essential endocytic function is indeed to stabilize membrane curvature. —SMH

J. Cell Biol. **221**, e202109013 (2022).

IMMUNOLOGY

Eosinophils busy in the background

Eosinophils are granulocytes that can be called into action during allergic reactions and immune responses to helminths, but what these cells do in tissues the rest of the time is poorly understood. Ignacio *et al.* report that eosinophils in the small intestine reside within the villous lamina propria, where they help to regulate the size of villi, macrophage maturation, and barrier integrity. This homeostatic behavior depends on the presence of gut microbiota, which stimulate intestinal epithelial cells to produce interleukin-33, a cytokine that regulates eosinophil activation and maturation. —STS

Immunity 10.1016/j.immuni.2022.05.014 (2022).

PHYSICS

Superconducting oddity

Odd-parity superconductors are predicted to have exotic properties, but the list of candidate materials is rather short. A recent addition to the list, $CeRh_2As_2$, has been found to harbor two distinct superconducting phases, one of which may be odd parity. Landaeta *et al.* undertook a systematic study of the properties of $CeRh_2As_2$, such as ac susceptibility, specific heat, and magnetic torque, while varying the angle of an applied magnetic field with respect to the *c*-axis. In conjunction with calculations, the measurements strengthen the evidence for the existence of an odd-parity superconducting state in this compound. —JS

Phys. Rev. X **12**, 031001 (2022).

PHOTO: TON ANURAK/SHUTTERSTOCK

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Immune boosting by B.1.1.529 (Omicron) depends on previous SARS-CoV-2 exposure

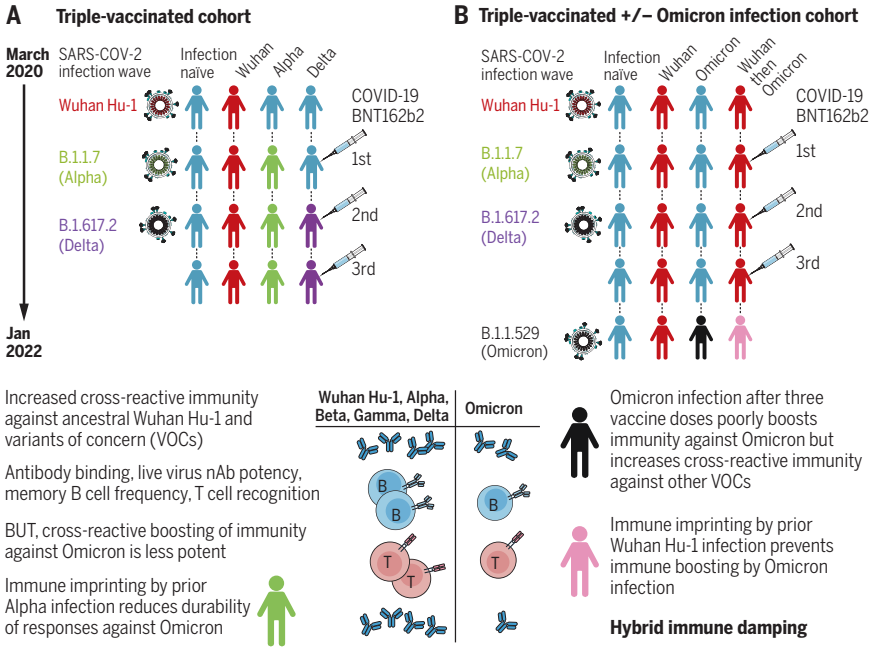
Catherine J. Reynolds†, Corinna Pade†, Joseph M. Gibbons†, Ashley D. Otter†, Kai-Min Lin, Diana Muñoz Sandoval, Franziska P. Pieper, David K. Butler, Siyi Liu, George Joy, Nasim Forooghi, Thomas A. Treibel, Charlotte Manisty, James C. Moon, COVIDsortium Investigators§, COVIDsortium Immune Correlates Network§, Amanda Semper, Tim Brooks, Áine McKnight‡, Daniel M. Altmann‡, Rosemary J. Boyton*‡

INTRODUCTION: B.1.1.529 (Omicron) and its subvariants pose new challenges for control of the COVID-19 pandemic. Although vaccinated populations are relatively protected from severe disease and death, countries with high vaccine uptake are experiencing substantial caseloads with breakthrough infection and frequent reinfection.

RATIONALE: We analyzed cross-protective immunity against B.1.1.529 (Omicron) in triple-vaccinated health care workers (HCWs) with different immune-imprinted histories of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection during the ancestral Wuhan Hu-1, B.1.1.7 (Alpha), and B.1.617.2 (Delta) waves and after infection during the B.1.1.529 (Omicron) wave in previously infection-naïve individuals and those with hybrid immunity, to investigate whether B.1.1.529 (Omicron) infection could further boost adaptive immunity. Spike subunit 1 (S1) receptor binding domain (RBD) and whole spike binding, live virus neutralizing antibody (nAb) potency, memory B cell (MBC) frequency, and T cell responses against peptide pools and naturally processed antigen were assessed.

RESULTS: B and T cell recognition and nAb potency were boosted against previous variants of concern (VOCs) in triple-vaccinated HCWs, but this enhanced immunity was attenuated against B.1.1.529 (Omicron) itself. Furthermore, immune imprinting after B.1.1.7 (Alpha) infection resulted in reduced durability of antibody binding against B.1.1.529 (Omicron), and S1 RBD and whole spike VOC binding correlated poorly with live virus nAb potency. Half of triple-vaccinated HCWs showed no T cell response to B.1.1.529 (Omicron) S1 processed antigen, and all showed reduced responses to the B.1.1.529 (Omicron) peptide pool, irrespective of SARS-CoV-2 infection history. Mapping T cell immunity in class II human leukocyte antigen transgenics showed that individual spike mutations could result in loss or gain of T cell epitope recognition, with changes to T cell effector and regulatory programs. Triple-vaccinated, previously infection-naïve individuals infected during the B.1.1.529 (Omicron) wave showed boosted cross-reactive S1 RBD and whole spike binding, live virus nAb potency, and T cell immunity against previous VOCs but less so against B.1.1.529 (Omicron) itself. Immune imprinting from prior Wuhan Hu-1 infection abrogated any enhanced cross-reactive antibody binding, T cell recognition, MBC frequency, or nAb potency after B.1.1.529 (Omicron) infection.

CONCLUSION: Vaccine boosting results in distinct, imprinted patterns of hybrid immunity with different combinations of SARS-CoV-2 infection and vaccination. Immune protection is boosted by B.1.1.529 (Omicron) infection in the triple-vaccinated, previously infection-naïve individuals, but this boosting is lost with prior Wuhan Hu-1 imprinting. This "hybrid immune damping" indicates substantial subversion of immune recognition and differential modulation through immune imprinting and may be the reason why the B.1.1.529 (Omicron) wave has been characterized by breakthrough infection and frequent reinfection with relatively preserved protection against severe disease in triple-vaccinated individuals. ■



Hybrid immune damping. (**A**) Triple-vaccinated HCWs with different SARS-CoV-2 infection histories show boosted cross-reactive immunity against VOCs, less so against Omicron. (**B**) Breakthrough infection during the Omicron wave boosts cross-reactive immunity in triple-vaccinated, previously infection-naïve individuals against VOCs, less so against Omicron itself; imprinting by previous Wuhan Hu-1 infection ablates Omicron immune boosting. (**C**) T cell recognition of Omicron mutation sequences is linked to altered transcription.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: r.boyton@imperial.ac.uk
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
§The members of the COVIDsortium Investigators and COVIDsortium Immune Correlates Network are listed in the supplementary materials.

Cite this article as C. J. Reynolds *et al.*, *Science* 377, eabq1841 (2022). DOI: 10.1126/science.abq1841

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abq1841

RESEARCH ARTICLE

CORONAVIRUS

Immune boosting by B.1.1.529 (Omicron) depends on previous SARS-CoV-2 exposure

Catherine J. Reynolds[1]†, Corinna Pade[2]†, Joseph M. Gibbons[2]†, Ashley D. Otter[3]†, Kai-Min Lin[1], Diana Muñoz Sandoval[1], Franziska P. Pieper[1], David K. Butler[1], Siyi Liu[1], George Joy[4], Nasim Forooghi[4], Thomas A. Treibel[4,5], Charlotte Manisty[4,5], James C. Moon[4,5], COVIDsortium Investigators§, COVIDsortium Immune Correlates Network§, Amanda Semper[3], Tim Brooks[3], Áine McKnight[2]‡, Daniel M. Altmann[6]‡, Rosemary J. Boyton[1,7]*‡

The Omicron, or Pango lineage B.1.1.529, variant of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) carries multiple spike mutations with high transmissibility and partial neutralizing antibody (nAb) escape. Vaccinated individuals show protection against severe disease, often attributed to primed cellular immunity. We investigated T and B cell immunity against B.1.1.529 in triple BioNTech BNT162b2 messenger RNA–vaccinated health care workers (HCWs) with different SARS-CoV-2 infection histories. B and T cell immunity against previous variants of concern was enhanced in triple-vaccinated individuals, but the magnitude of T and B cell responses against B.1.1.529 spike protein was reduced. Immune imprinting by infection with the earlier B.1.1.7 (Alpha) variant resulted in less durable binding antibody against B.1.1.529. Previously infection-naïve HCWs who became infected during the B.1.1.529 wave showed enhanced immunity against earlier variants but reduced nAb potency and T cell responses against B.1.1.529 itself. Previous Wuhan Hu-1 infection abrogated T cell recognition and any enhanced cross-reactive neutralizing immunity on infection with B.1.1.529.

At the end of 2021, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron variant of concern (VOC) spread rapidly, displacing the prior most prevalent VOC, B.1.617.2 (Delta) (*1*, *2*). B.1.1.529 (Omicron) diverges more from the ancestral Wuhan Hu-1 sequence than other VOCs so far, with 36 coding mutations in the spike protein, associated with high transmission, tendency to infect cells of the upper bronchus, and presentation with flu-like symptoms (*3*–*5*). Across several studies, two- or three-dose vaccination is protective against severe disease and hospitalization, albeit with poor protection against transmission (*6*–*8*). A rationale for this high rate of breakthrough infections comes from mapping of virus neutralization using either postvaccination immune sera or monoclonal antibodies, showing Omicron to be the most antibody immune-evasive VOC, with titers generally reduced 20- to 40-fold (*9*–*12*). The relative attenuation of severe symptoms in vaccinated compared with unvaccinated groups is likely attributable to the partial protection conferred by the residual neutralizing antibody (nAb) repertoire and the activation of primed B cell and T cell memory (*13*–*18*). In this study, we applied our ongoing analysis of a cohort of London health care workers (HCWs) (*19*–*24*) to address two key issues of B.1.1.529 (Omicron) immunity. First, following the earlier demonstration that people at this stage in the pandemic carry heterogeneous, immune-imprinted repertoires derived from their distinctive histories of infection and vaccination, we explored how these differences manifest in differential cross-recognition of B.1.1.529 (Omicron) relative to other VOCs, at the level of binding and neutralizing Ab, B cell, and T cell immunity (*24*). Analyzing a London HCW cohort having detailed longitudinal, clinical, transcriptomic, and immunologic characterization, we considered the extent to which prior encounter with spike antigen through infection and vaccination shapes subsequent immunity to B.1.1.529 (Omicron) through immune imprinting. Second, when B.1.1.529 (Omicron) infections and reinfections have been so pervasive (*25*), it is possible that B.1.1.529 (Omicron) infection may confer a benign, live booster to vaccine immunity. Hence, we investigated the extent to which B.1.1.529 (Omicron) infection boosts cross-reactive B and T cell immunity against other VOCs and itself.

[1]Department of Infectious Disease, Imperial College London, London, UK. [2]Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [3]UK Health Security Agency, Porton Down, UK. [4]St Bartholomew's Hospital, Barts Health NHS Trust, London, UK. [5]Institute of Cardiovascular Science, University College London, London, UK. [6]Department of Immunology and Inflammation, Imperial College London, London, UK. [7]Lung Division, Royal Brompton and Harefield Hospitals, Guy's and St Thomas' NHS Foundation Trust, London, UK.
*Corresponding author. Email: r.boyton@imperial.ac.uk
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
§The members of the COVIDsortium Investigators and COVIDsortium Immune Correlates Network are listed in the supplementary materials.

Results

B cell immunity after three vaccine doses

A London cohort of HCWs were followed longitudinally from March 2020 to January 2022. We identified HCWs with mild and asymptomatic SARS-CoV-2 infection by ancestral Wuhan Hu-1, B.1.1.7 (Alpha VOC), B.1.617.2 (Delta VOC), and then B.1.1.529 (Omicron VOC) during successive waves of infection and after first, second, and third mRNA (BioNTech BNT162b2) vaccine doses (Fig. 1A, fig. S1, and table S1). We identified individuals with different combinations of SARS-CoV-2 infection and vaccination histories to study the impact of immune imprinting. Nucleocapsid (N) and subunit 1 (S1) spike receptor binding domain (RBD) serology was monitored longitudinally (Fig. 1B). As previously reported, the third spike exposure boosted the majority of HCWs above an S1 RBD titer of 1/10,000 binding antibody units per milliliter (U/ml) at 2 to 3 weeks after the most recent vaccine dose. By three vaccine doses, antibody responses had plateaued, regardless of infection history (*24*).

In triple-vaccinated HCWs 2 to 3 weeks after their third dose (table S1), we compared antibody titers against RBD (Fig. 1C), whole spike (fig. S2), and live virus nAb half-maximal inhibitory concentration (IC_{50}) (Fig. 1D) for ancestral Wuhan Hu-1 and each of the VOCs (table S2). We stratified the vaccinated HCWs according to whether they were infection-naïve or had previously been infected with Wuhan Hu-1, B.1.1.7 (Alpha), or B.1.617.2 (Delta) (Fig. 1A). We found differences in immune imprinting indicating that those who were infected during the ancestral Wuhan Hu-1 wave showed a significantly reduced anti-RBD titer against B.1.351 (Beta), P.1 (Gamma), and B.1.1.529 (Omicron) compared with infection-naïve HCWs (Fig. 1C). The hybrid immune groups that had experienced previous Wuhan Hu-1 and B.1.1.7 (Alpha) infection showed more potent nAb responses against Wuhan Hu-1, B.1.1.7 (Alpha), and B.1.617.2 (Delta) (Fig. 1D). However, cross-reactive S1 RBD immunoglobulin G (IgG) antibody and nAb IC_{50} against B.1.1.529 (Omicron) were significantly reduced compared with the other VOCs, irrespective of previous SARS-CoV-2 infection history (table S3 and Fig. 1, C and D).

Memory B cell (MBC) frequencies against ancestral Wuhan Hu-1, B.1.617.2 (Delta), and B.1.1.529 (Omicron) S1 protein were boosted 2 to 3 weeks after the third vaccine dose compared with 20 to 21 weeks after the second vaccine dose (Fig. 1E). Irrespective of infection history, the MBC frequencies against Wuhan Hu-1 and B.1.617.2 (Delta) S1 were similar, but those against B.1.1.529 (Omicron) S1 were significantly reduced 2 to 3 weeks after the third vaccine dose [reduction in B.1.1.529 (Omicron) S1 MBC frequency compared with Wuhan Hu-1 S1 was 2.0-fold (P = 0.0156) for infection-naïve, 2.4-fold for Wuhan Hu-1 infection (P = 0.0020),

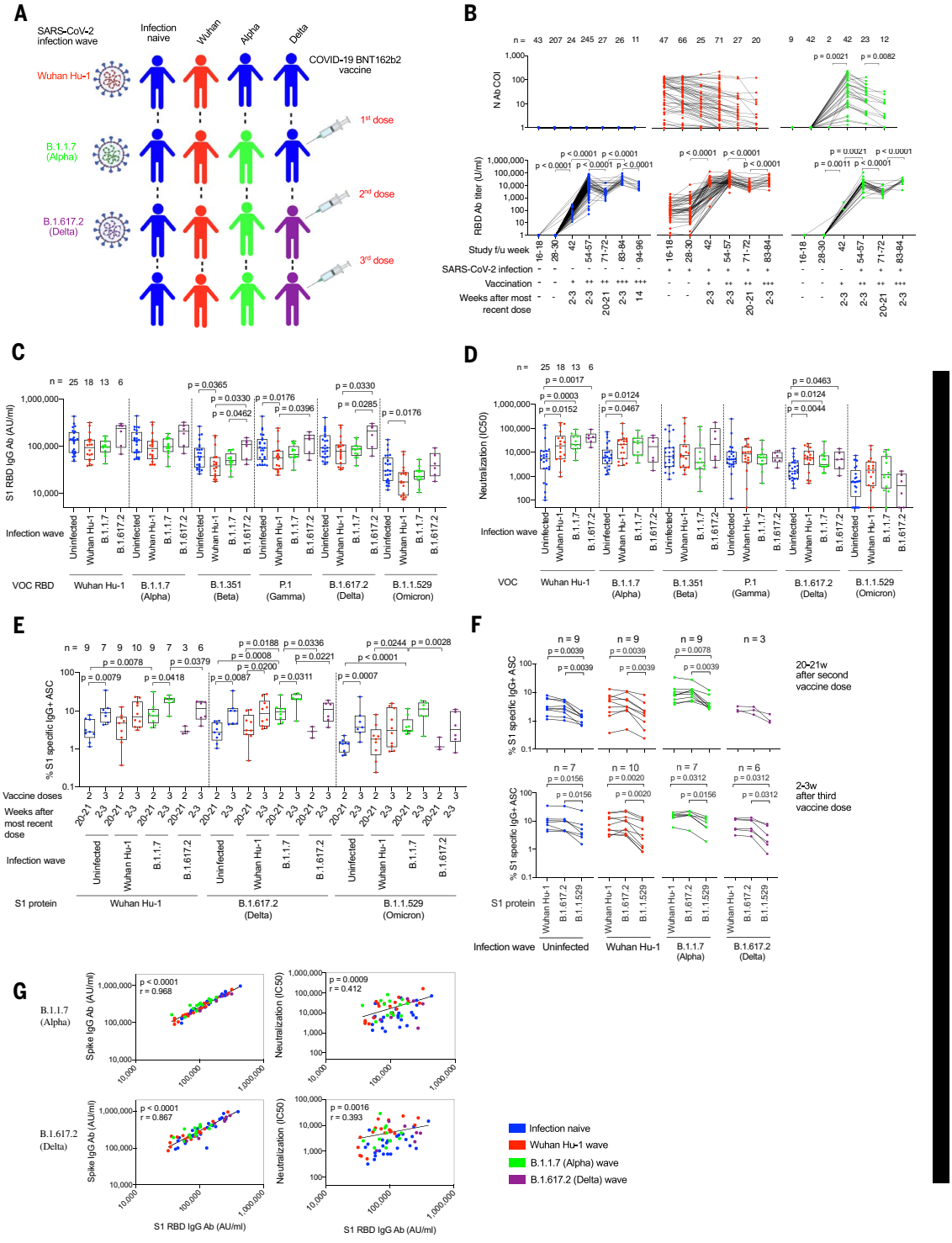


Fig. 1. SARS-CoV-2 infection history alters Ab and B cell immunity in triple-vaccinated HCWs. (**A**) Graphical summary depicting the SARS-CoV-2 infection and vaccination history of HCWs studied. (**B**) Serum Ab binding against SARS-CoV-2 N (top panel) and ancestral Wuhan Hu-1 S1 RBD (bottom panel) in infection-naïve HCWs (blue, n = 11 to 245) and HCWs with laboratory-confirmed SARS-CoV-2 infection during the Wuhan Hu-1 (red, n = 20 to 71) or B.1.1.7 (Alpha, green, n = 12 to 42) waves. Data are shown pre-vaccination and at defined time points after a first, second, and third dose of BNT162b2. (**C**) Serum S1 RBD Ab binding and (**D**) nAb IC_{50} against ancestral Wuhan Hu-1, B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), and B.1.1.529 (Omicron) live virus 2 to 3 weeks after the third vaccine dose in infection-naïve HCWs (blue, n = 25) or HCWs with laboratory-confirmed SARS-CoV-2 infection during the ancestral Wuhan Hu-1 (red, n = 18), B.1.1.7 (Alpha, green, n = 13), or B.1.617.2 (Delta, purple, n = 6) waves. (**E**) Frequency of MBC specific for ancestral Wuhan Hu-1, B.1.617.2 (Delta), and B.1.1.529 (Omicron) spike S1 protein 20 to 21 weeks after the second and 2 to 3 weeks after the third vaccine dose in infection-naïve (blue, n = 7 to 9) or HCWs infected by SARS-CoV-2 during the Wuhan Hu-1 (red, n = 9 to 10), B.1.1.7 (Alpha, green, n = 7 to 9), or B.1.617.2 (Delta, purple, n = 3 to 6) waves. (**F**) MBC frequency data plotted pairwise for individual HCWs at 20 to 21 weeks after second dose (top panel) or 2 to 3 weeks after third dose (bottom panel). (**G**) Correlations between S1 RBD VOC and whole spike VOC Ab binding (left-hand panel) and nAb IC_{50} (right-hand panel) against B.1.1.7 (Alpha) and B.1.617.2 (Delta) in infection-naïve (blue, n = 25) HCWs and HCWs infected during the ancestral Wuhan Hu-1 (red, n = 18), B.1.1.7 (Alpha, green, n = 13), and B.1.617.2 (Delta, purple, n = 6) waves. Statistical tests were performed using Prism 9.0. [(B) to (E)] Mann-Whitney U test, (F) Wilcoxon matched-pairs signed rank test, (G) Spearman's rank correlation. f/u; follow-up; w, weeks.

1.9-fold for B.1.1.7 (Alpha) infection (P = 0.0312), and 2.9-fold for B.1.617.2 (Delta) infection (P = 0.0312)] and at 20 to 21 weeks after the second dose [reduction in B.1.1.529 (Omicron) MBC frequency compared with Wuhan Hu-1 was 2.5-fold (P = 0.0039) for infection-naïve HCWs and 2.2-fold (P = 0.0039), twofold (P = 0.0078), and 2.9-fold (P = 0.1250) for Wuhan Hu-1, B.1.1.7 (Alpha), and B.1.617.2 (Delta) infection groups, respectively] (Fig. 1F and table S4).

S1 RBD or whole spike antibody binding and live virus nAb IC_{50} correlated for B.1.1.7 (Alpha) and B.1.617.2 (Delta), but not for B.1.351 (Beta), P.1 (Gamma), or B.1.1.529 (Omicron), indicating that antibody binding serology was a poor marker of nAb IC_{50} (Fig. 1G and fig. S3). Differences between VOC RBD and whole spike binding and nAb IC_{50} with live virus indicated

that antibody targeting regions outside RBD and/or spike or conformational epitopes exposed only during infection may contribute to neutralization (*26*, *27*) (Fig. 1, C and D).

T cell immunity after three vaccine doses

We next compared T cell responses at 2 to 3 weeks after the third dose in triple-vaccinated HCWs who were either infection-naïve or had been infected during the Wuhan Hu-1, B.1.1.7 (Alpha), or B.1.617.2 (Delta) wave (fig. S1 and table S1). We compared T cell immunity against a mapped epitope pool (MEP) of ancestral Wuhan Hu-1 spike peptides (table S5A) with spike S1 protein from ancestral Wuhan Hu-1 or S1 proteins containing the B.1.617.2 (Delta) or B.1.1.529 (Omicron) mutations. Measuring T cell responses against naturally processed epitopes from whole S1 protein antigen of ancestral Wuhan Hu-1, B.1.617.2 (Delta), or B.1.1.529 (Omicron) sequence allowed us to focus on immunodominant responses representative of those presented in real-life infection. T cell responses against Wuhan Hu-1 S1 protein mirrored those elicited by MEP stimulation, with the majority making a strong response (Fig. 2A). However, for S1 B.1.1.529 (Omicron) protein, we found a significantly reduced magnitude of response. Overall, more than half (27/50, 54%) made no T cell response against S1 B.1.1.529 (Omicron) protein, irrespective of previous SARS-CoV-2 infection history, compared with 8% (4/50) that made no T cell response against ancestral Wuhan Hu-1 S1 protein ($P < 0.0001$, Chi-square test) (Fig. 2A). The fold-reduction in geometric mean of T cell response [spot forming cells (SFCs)] against B.1.1.529 (Omicron) S1 compared with ancestral Wuhan Hu-1 S1 protein was 17.3-fold for infection-naïve HCWs (blue, $P < 0.0001$), 7.7-fold for previously Wuhan Hu-1 infected (red, $P = 0.001$), 8.5-fold for previously B.1.1.7 (Alpha) infected (green, $P = 0.002$), and 19-fold for previously B.1.617.2 (Delta) infected (purple, $P = 0.0625$) HCWs (Fig. 2B).

To investigate T cell recognition of VOC sequence mutations, we used a peptide pool specifically designed to cover all of the B.1.1.529 (Omicron) S1 and S2 spike mutations and a matched pool containing the equivalent Wuhan Hu-1 sequences (Fig. 2C and table S5B). T cell responses against the B.1.1.529 (Omicron) peptide pool were reduced compared with the Wuhan Hu-1 pool, irrespective of previous infection history [reduction in T cell response against B.1.1.529 (Omicron) peptide pool compared with equivalent ancestral Wuhan Hu-1 peptide pool was 2.7-fold for infection-naïve (blue, $P = 0.0003$), 4.6-fold for previously Wuhan Hu-1 infected (red, $P = 0.0039$), 2.7-fold for previously B.1.1.7 (Alpha) infected (green, $P = 0.0039$), and 3.8-fold for previously B.1617.2 (Delta) infected HCWs (purple, $P = 0.1250$)] (Fig. 2C). In fact, 42% (21/50) of HCWs made no T cell response at all against the B.1.1.529 (Omicron) VOC mutant pool (Fig. 2C).

Overall, our findings in triple-vaccinated HCWs with different previous SARS-CoV-2 infection histories indicated that T cell cross-recognition of B.1.1.529 (Omicron) S1 antigen and peptide pool was significantly reduced. T cell and nAb responses against B.1.1.529

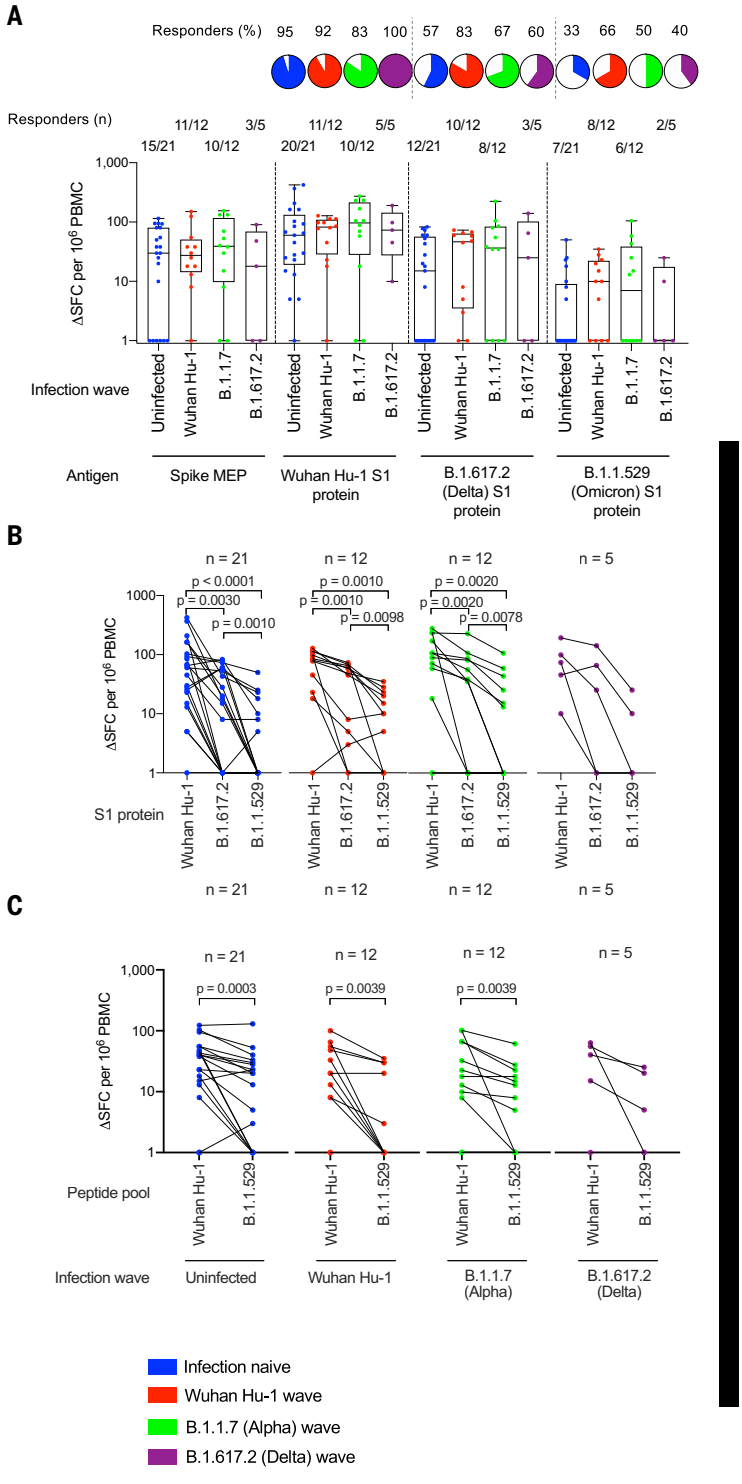


Fig. 2. T cell cross-recognition of B.1.1.529 (Omicron) in triple-vaccinated HCWs. (**A**) T cell responses against ancestral Wuhan Hu-1 spike MEP pool or ancestral Wuhan Hu-1, B.1.617.2 (Delta), and B.1.1.529 (Omicron) VOC S1 proteins in PBMC from infection-naïve HCWs (blue) or HCWs with laboratory-confirmed SARS-CoV-2 infection during the ancestral Wuhan Hu-1 (red), B.1.1.7 (Alpha, green), and B.1.617.2 (Delta, purple) waves. PBMCs were taken 2 to 3 weeks after the third vaccine dose, and T cell responses assessed by IFNγ ELISpot. Pie charts show the percent of responders with a detectable T cell response against each antigen. (**B**) Spike MEP pool and S1 protein T cell responses plotted pair-wise for each individual HCW. (**C**) T cell responses against peptide pools containing either the B.1.1.529 (Omicron) mutations found in the SARS-CoV-2 spike or the equivalent original ancestral Wuhan Hu-1 sequences. PBMCs from infection-naïve HCWs (blue) or HCWs infected during the ancestral Wuhan Hu-1 (red), B.1.1.7 (Alpha, green), or B.1.617.2 (Delta, purple) waves were stimulated by peptide pools containing the original Wuhan Hu-1 or B.1.1.529 sequences and plotted pair-wise. Statistical tests were performed using Prism 9.0. (A) Mann-Whitney *U* test, [(B) and (C)] Wilcoxon matched-pairs signed rank test.

(Omicron) were discordant, and most (20/27, 74%) HCWs with no T cell response against B.1.1.529 (Omicron) S1 made cross-reactive nAb against B.1.1.529 at an IC_{50} of >195 (fig. S4).

B.1.1.529 (Omicron) spike mutations encompass gain and loss of T cell epitopes

Owing to the complexities inherent in mapping the effects of mutations in individual T cell epitopes across cohorts carrying heterogeneous human leukocyte antigen (HLA) alleles, we mapped the differential recognition of the B.1.1.529 (Omicron) spike mutations using HLA-DRB1*04:01 transgenic mice (*23*, *24*) (Fig. 3). The peptide pool containing B.1.1.529

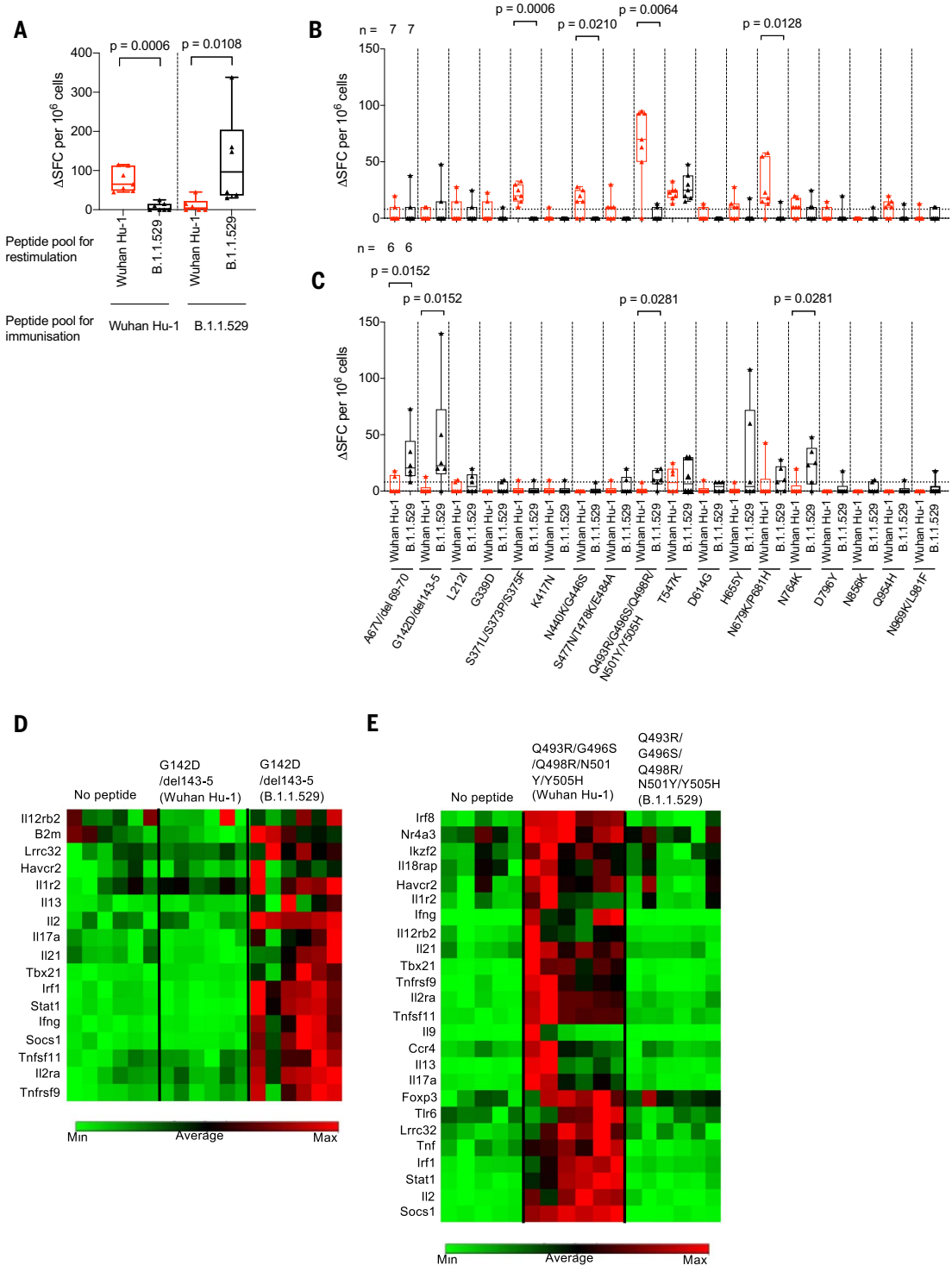


Fig. 3. B.1.1.529 (Omicron) spike mutations alter T cell recognition. (**A**) HLAII transgenic mice carrying DRB1*0401 in the context of a homozygous knockout for murine H2-Aβ (7 to 8 weeks old) were immunized with either a B.1.1.529 (Omicron, *n* = 7) VOC pool of 18 peptides encompassing the Omicron sequence mutations or the ancestral Wuhan Hu-1 pool of peptides (*n* = 6) with the equivalent unmutated sequences. At day 10, draining lymph node (DLN) cells were prepared from immunized mice and stimulated with either Wuhan Hu-1 (red) or B.1.1.529 (Omicron, black) peptide pools, and T cell responses were measured by IFNγ ELISpot. (**B**) IFNγ T cell responses were mapped against individual Wuhan Hu-1 (red) or B.1.1.529 (Omicron, black) peptides using DLN cells taken from Wuhan Hu-1 peptide pool (*n* = 7) or (**C**) B.1.1.529 (Omicron) peptide pool (*n* = 6) immunized mice. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**D** and **E**) Heatmaps showing relative gene expression of T cell activation markers in DLN cells taken from B.1.1.529 (Omicron) G142D/del143-5 peptide primed (D) (*n* = 6) or Wuhan Hu-1 Q493R/G496S/Q498R/N501Y/Y505H peptide primed (E) (*n* = 6) HLA-DRB1*04:01 transgenic mice. DLN cells were stimulated for 24 hours in vitro with 10 μg/ml Wuhan Hu-1 or B.1.1.529 (Omicron) variant peptide before RNA extraction. Genes shown in the heatmap are significantly up-regulated ($P < 0.05$) in Wuhan Hu-1 or B.1.1.529 (Omicron) variant peptide stimulated cells compared with no peptide control. Statistical tests were performed using Prism 9.0 or the Qiagen GeneGlobe data analysis tool for gene expression data. [(A) to (C)] Wilcoxon matched-pairs signed rank test, [(D) and (E)] Student's *t* test.

(Omicron)–specific S1 and S2 spike mutations and its ancestral Wuhan Hu-1 equivalent pool showed differential, sequence-specific T cell priming by either ancestral Wuhan Hu-1 or B.1.1.529 (Omicron) sequence-specific peptide pools (Fig. 3A and table S5B). That is, immunizing HLAII transgenic mice with either ancestral Wuhan Hu-1 or B.1.1.529 (Omicron) sequence-specific peptide pools allowed us to investigate differential, sequence-specific T cell priming that occurs as a consequence of B.1.1.529 (Omicron) spike mutations. We showed that priming with one pool resulted in impaired responses to the other (Fig. 3A). We then looked at responses to individual HLA-DRB1*04:01 epitopes. Notably, while the B.1.1.529 (Omicron) mutations were associated in four instances with loss of a clear HLA-DR4–restricted T cell epitope (Fig. 3B: S371L/S373P/S375F, $P = 0.0006$; N440K/G446S, $P = 0.0210$; Q493R/G496S/Q498R/N501Y/Y505H, $P = 0.0064$; N679K/P681H, $P = 0.0128$), the mutated sequence epitopes in four instances led to de novo gain of Omicron-specific HLA-DR4 T cell epitopes (Fig. 3C: A67V/del69-70, $P = 0.0152$; G142D/del143-5, $P = 0.0152$; Q493R/G496S/Q498R/N501Y/Y505H/N679K/P681H, $P = 0.0281$; N764K, $P = 0.0281$). The G142D/del143-5 mutation created a gain-of-function epitope, switching from a region not recognized by T cells to one that induced a T helper 1 (T_H1)/T_H17 effector program (Fig. 3D). We have previously shown that the N501Y mutation converts a T cell effector–stimulating epitope to an inducer of immune regulation (*24*). This finding was confirmed and reiterated by the more extensive alterations in the Q493R/G496S/Q498R/N501Y/Y505H mutant epitope (Fig. 3E).

B cell immunity after B.1.1.529 (Omicron) infection

Next, we studied triple-vaccinated HCWs 14 weeks after their third dose, who had suffered breakthrough infection during the B.1.1.529 (Omicron) wave. These individuals were compared with infection-naïve and previously Wuhan Hu-1 infected HCWs who had escaped B.1.1.529 (Omicron) wave infection (Fig. 4A, tables S6 and S7, and fig. S1). Previously Wuhan Hu-1 infected HCWs who were also infected during the B.1.1.529 (Omicron) wave showed the highest N antibody binding (Fig. 4B and tables S6 and S7). Previously infection-naïve triple-vaccinated HCWs made significantly increased cross-reactive antibody binding responses against all VOCs and B.1.1.529 (Omicron) itself after infection during the B.1.1.5129 (Omicron) wave: S1 RBD (Fig. 4C and table S2), whole spike (Fig. 4D and table S2), and nAb IC_{50} (Fig. 4E). However, antibody binding and nAb IC_{50} were attenuated against B.1.1.529 (Omicron) itself with a 4.5-fold reduction in S1 RBD binding ($P = 0.001$) and a 10.1-fold reduction in nAb IC_{50} ($P = 0.002$) against B.1.1.529 compared with ancestral Wuhan Hu-1. Thus, infection during the B.1.1.529 (Omicron) wave produced potent cross-reactive antibody immunity against all VOCs, but less so against B.1.1.529 (Omicron) itself.

Triple-vaccinated, infection-naïve HCWs who were not infected during the B.1.1.529 (Omicron) wave made no nAb IC_{50} response against B.1.1.529 (Omicron) 14 weeks after the third vaccine dose, indicating rapid waning of the nAb IC_{50} from a mean value of 1400 at 2 to 3 weeks to 0 at 14 weeks after the third dose ($P = 0.0312$) (Fig. 4E and fig. S5A).

HCWs with a history of prior Wuhan Hu-1 infection who were also infected during the B.1.1.529 (Omicron) wave showed no increase in cross-reactive S1 RBD (Fig. 4C) or whole spike (Fig. 4D) antibody binding or live virus nAb IC_{50} (Fig. 4E and fig. S5B) against B.1.1.529 (Omicron) or any other VOC, despite having made a higher N antibody response (Fig. 4B). Thus, B.1.1.529 (Omicron) infection can boost binding and nAb responses against itself and other VOCs in triple-vaccinated previously uninfected infection-naïve HCWs but not in the context of immune imprinting after prior Wuhan Hu-1 infection. Immune imprinting by prior Wuhan Hu-1 infection completely abrogated any enhanced nAb responses against B.1.1.529 (Omicron) and other VOCs (Fig. 4E).

Increased MBC frequency against ancestral Wuhan Hu-1, B.1.617.2 (Delta), and B.1.1.529 (Omicron) S1 and RBD proteins was observed in previously infection-naïve HCWs infected during the B.1.1.529 (Omicron) wave (Fig. 4F). This was not the case for HCWs who had been previously infected during the first Wuhan Hu-1 wave and then infected again during the B.1.1.529 (Omicron) wave (Fig. 4F).

In summary, B.1.1.529 (Omicron) infection resulted in enhanced, cross-reactive Ab responses against all VOCs tested in the three-dose vaccinated infection-naïve HCWs but not in those with previous Wuhan Hu-1 infection, and less so against B.1.1.529 (Omicron) itself (Fig. 4, C to E). In line with this, MBC frequencies against Wuhan Hu-1, B.1.617.2 (Delta), and B.1.1.529 (Omicron) S1 and RBD proteins increased in three-dose vaccinated, infection-naïve individuals but not in those imprinted by previous Wuhan Hu-1 infection (Fig. 4F).

S1 RBD or whole spike antibody binding and live virus nAb IC_{50} correlated for B.1.1.529 (Omicron) [correlation coefficient (r) of 0.687, $P < 0.0001$] and all the VOCs tested ($r > 0.539$), but there was considerable discordance in that many of the HCWs recording no detectable live virus nAb IC_{50} against B.1.1.529 (Omicron) recorded S1 RBD (Omicron) binding serology ranging from 3412 to 20,484, indicating that S1 RBD (VOC) antibody binding serology could be misleading and a poor marker of nAb (Fig. 4G and fig. S6).

T cell immunity after B.1.1.529 (Omicron) infection

We next explored T cell immunity after breakthrough infection during the B.1.1.529 (Omicron) wave. Fourteen weeks after the third dose (9/10, 90%) of triple-vaccinated, previously infection-naïve HCWs showed no cross-reactive T cell immunity against B.1.1.529 (Omicron) S1 protein (Fig. 5A).

The T cell response against B.1.1.529 (Omicron) S1 protein after infection during the B.1.1.529 (Omicron) wave of previously infection-naïve HCWs was significantly reduced compared with Wuhan Hu-1 S1 and B.1.617.2 (Delta) S1 in triple-vaccinated HCWs [geometric mean: 57, 50, and 6 SFCs for Wuhan Hu-1, B.1.617.2 (Delta), and B.1.1.529 (Omicron) S1 proteins, respectively; $P = 0.001$)] (Fig. 5B). HCWs infected during the B.1.1.529 (Omicron) wave showed similar T cell responses against spike MEP, ancestral Wuhan Hu-1 S1, and B.1.617.2 (Delta) S1 proteins but significantly reduced T cell responses against B.1.1.529 (Omicron) S1 protein [reduction in geometric mean of T cell response (SFCs) against VOC S1 protein compared with that for Wuhan Hu-1 S1: 1.1-fold reduction for B.1.617.2 S1, $P = 0.6836$; 10-fold reduction for B.1.1.529 S1, $P = 0.001$] (Fig. 5B). Thus, although breakthrough infection in triple-vaccinated HCWs during the Omicron infection wave boosted cross-reactive T cell immune recognition against the spike MEP pool ($P = 0.0117$), ancestral Wuhan Hu-1 ($P = 0.0039$), B.1.617.2 (Delta) ($P = 0.0003$), and B.1.1.529 (Omicron) (Fig. 5A), the T cell response against B.1.1.529 (Omicron) S1 protein itself compared with spike MEP ($P = 0.001$), Wuhan Hu-1 ($P = 0.001$), and B.1.617.2 (Delta) ($P = 0.001$) was significantly reduced (Fig. 5, B and C).

Notably, none (0/6) of HCWs with a previous history of SARS-CoV-2 infection during the Wuhan Hu-1 wave responded to B.1.1.529 (Omicron) S1 protein (Fig. 5A). This suggests that, in this context, B.1.1.529 (Omicron) infection was unable to boost T cell immunity against B.1.1.529 (Omicron) itself; immune imprinting from prior Wuhan Hu-1 infection resulted in the absence of a T cell response against B.1.1.529 (Omicron) S1 protein. These findings were further highlighted in paired data showing the fall in T cell response in individual HCWs across the three antigens: On an individual basis, most HCWs retained T cell recognition of B.1.617.2 S1 but commonly showed impaired T cell recognition of B.1.1.529 S1 (Fig. 5C). Taken together with the data shown in Fig. 4, the findings consistently demonstrate that people initially infected by Wuhan Hu-1 in the first wave and then reinfected during the B.1.1.529 (Omicron) wave do not experience a boost in T cell immunity against B.1.1.529 (Omicron) at the level of nAb and T cell recognition.

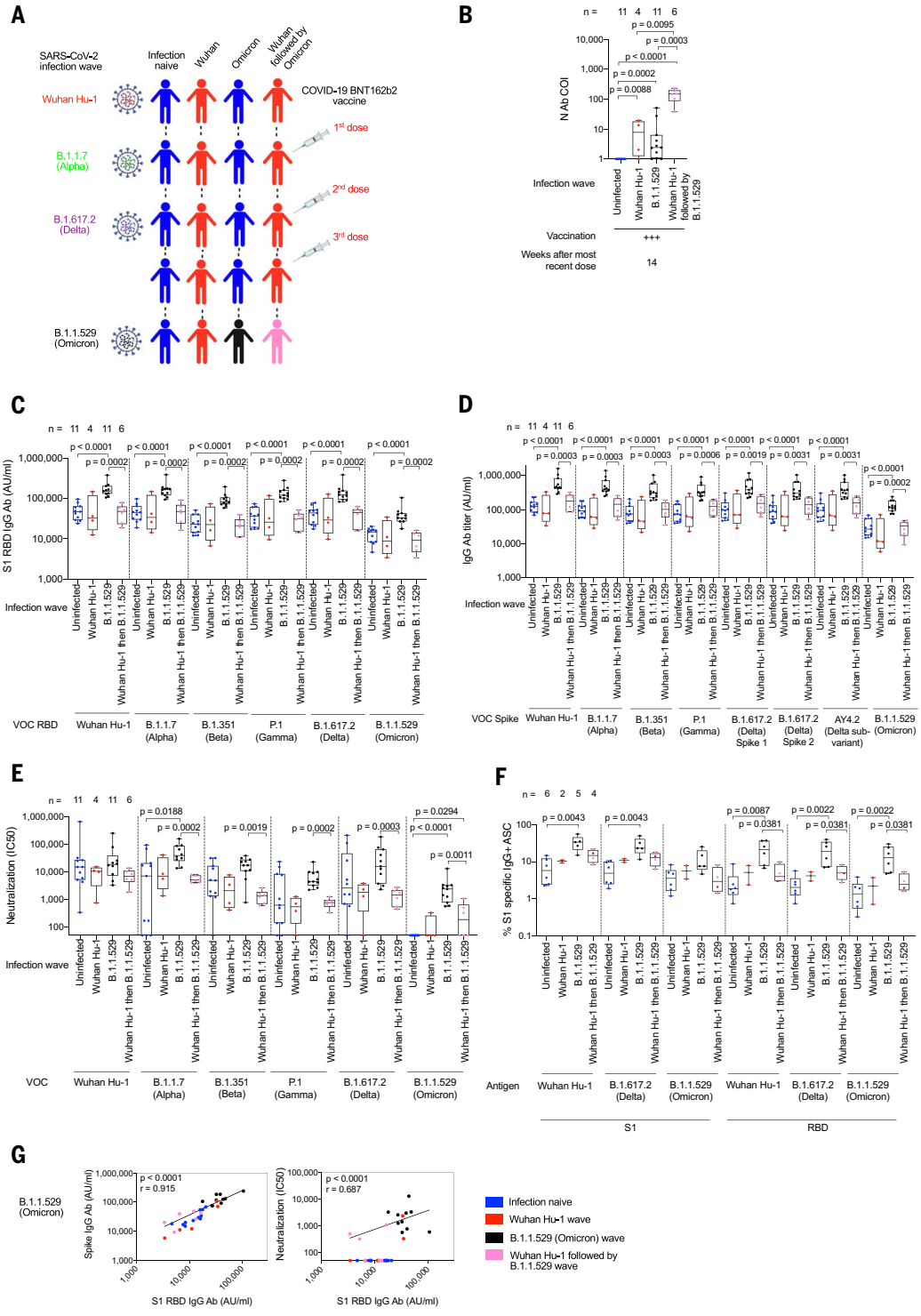


Fig. 4. Ab and B cell immunity in triple-vaccinated HCWs after infection during the B.1.1.529 (Omicron) wave. (**A**) Graphical summary depicting the SARS-CoV-2 infection and vaccination history of HCWs studied during the B.1.1.529 (Omicron) wave. (**B**) Serum Ab binding against SARS-CoV-2 N at 14 weeks (median, 14 weeks; IQR, 3 weeks) after third vaccine dose in infection-naïve HCWs (blue, *n* = 11) or in HCWs with laboratory-confirmed SARS-CoV-2 infection during the ancestral Wuhan Hu-1 (red, *n* = 4), B.1.1.529 (Omicron, black, *n* = 11), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, *n* = 6) infection waves. (**C**) Serum S1 RBD (VOC) Ab binding against ancestral Wuhan Hu-1, B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), and B.1.1.529 (Omicron) proteins in infection-naïve HCWs (blue, *n* = 11) or HCWs previously infected during the ancestral Wuhan Hu-1 (red, *n* = 4), B.1.1.529 (Omicron, black, *n* = 11), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, *n* = 6) waves. (**D**) Serum Ab binding against ancestral Wuhan Hu-1, B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), AY4.2 (Delta subvariant), and B.1.1.529 (Omicron) whole spike proteins in infection-naïve HCWs (blue, *n* = 11) or HCWs previously infected during the ancestral Wuhan Hu-1 (red, *n* = 4), B.1.1.529 (Omicron, black, *n* = 11), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, *n* = 6) waves. (**E**) Neutralizing antibody IC_{50} against Wuhan Hu-1 or VOC live virus isolates in infection-naïve HCWs (blue, *n* = 11) or HCWs previously infected during the ancestral Wuhan Hu-1 (red, *n* = 4), B.1.1.529 (Omicron, black, *n* = 11), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, *n* = 6) waves. (**F**) Frequency of MBC specific for ancestral Wuhan Hu-1, B.1.617.2 (Delta), and B.1.1.529 (Omicron) S1 and RBD binding proteins in infection-naïve HCWs (blue, *n* = 11) or HCWs previously infected during the ancestral Wuhan Hu-1 (red, *n* = 4), B.1.1.529 (Omicron, black, *n* = 11), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, *n* = 6) waves. (**G**) Correlation between whole spike and S1 RBD Ab binding (left-hand panel) or nAb IC_{50} and S1 RBD Ab binding (right-hand panel) for B.1.1.529 (Omicron) VOC in infection-naïve (blue, *n* = 11) or HCWs infected during the ancestral Wuhan Hu-1 (red, *n* = 4), B.1.1.529 (Omicron, black, *n* = 11), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, *n* = 6) waves. All data shown are from samples taken at 14 weeks (median, 14 weeks; IQR, 3 weeks) after third vaccine dose. Statistical tests were performed using Prism 9.0. [(B) to (F)] Mann-Whitney *U* test, (G) Spearman's rank correlation.

Prior infection differentially imprints Omicron T and B cell immunity

To investigate in more detail the impact of prior SARS-CoV-2 infection on immune imprinting, we further explored responses in our longitudinal HCW cohort (Fig. 6A and fig. S1). We first looked at the S1 RBD (ancestral Wuhan Hu-1 and Omicron VOC) antibody binding responses across the longitudinal cohort at key vaccination and SARS-CoV-2 infection time points, exploring how different exposure imprinted differential cross-reactive immunity and durability. This revealed that at 16 to 18 weeks after Wuhan Hu-1 or B.1.1.7 (Alpha) infection, unvaccinated HCWs showed no detectable

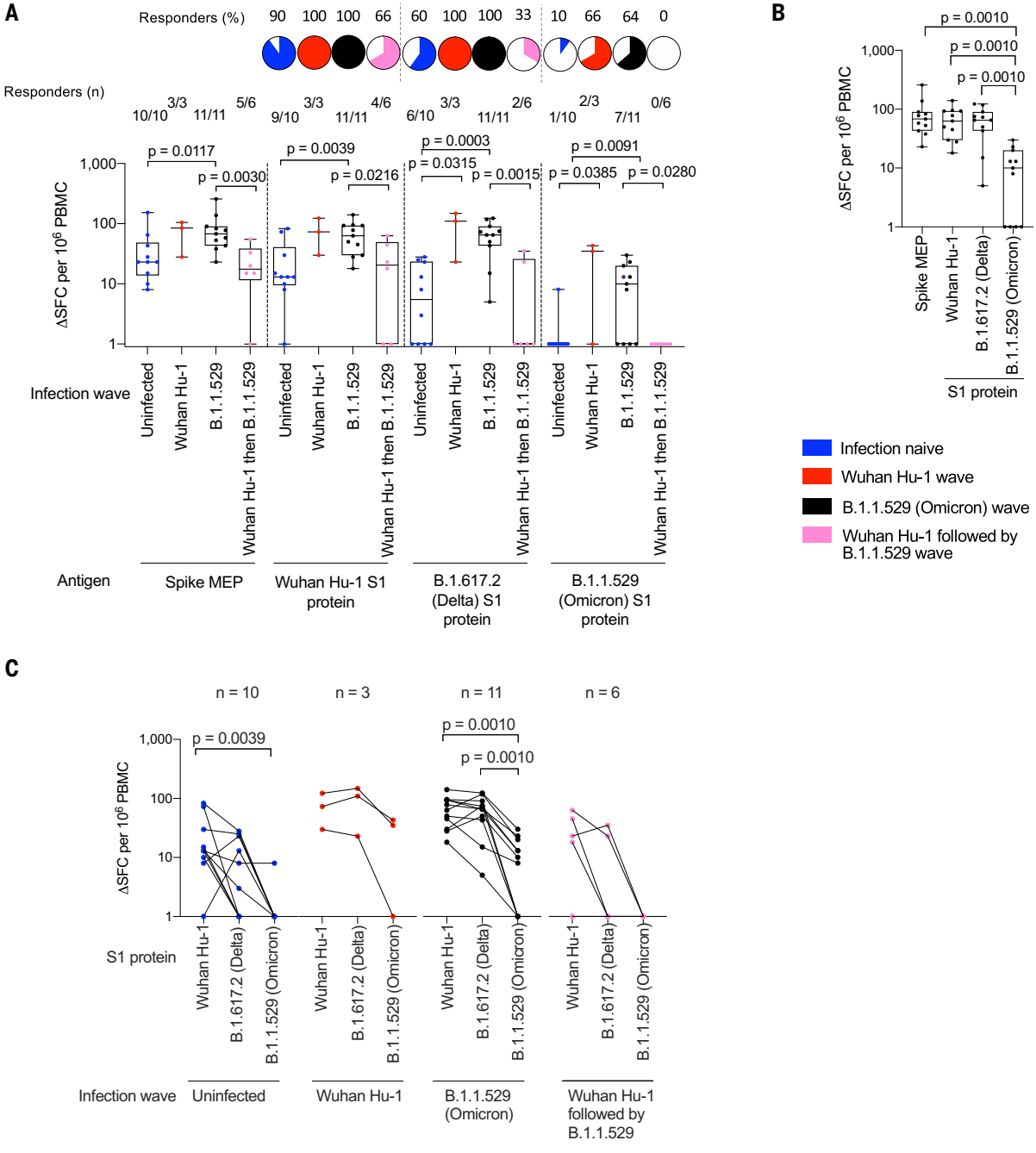


Fig. 5. T cell responses in triple-vaccinated HCWs infected during the B.1.1.529 (Omicron) wave. (**A**) T cell responses against ancestral Wuhan Hu-1 spike MEP pool or Wuhan Hu-1, B.1.617.2 (Delta), and B.1.1.529 (Omicron) VOC S1 protein for PBMCs taken from infection-naïve HCWs (blue, *n* = 10) or HCWs with laboratory-confirmed SARS-CoV-2 infection during the Wuhan Hu-1 (red, *n* = 3), B.1.1.529 (Omicron, black, *n* = 11), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, *n* = 6) waves. PBMCs were taken 14 weeks (median, 14 weeks; IQR, 3 weeks) after the third vaccine dose, and T cell responses were assessed by IFNγ ELISpot. Pie charts show the percent who had a detectable T cell response against each antigen. (**B**) T cell responses against spike MEP pool and S1 VOC protein for previously infection-naïve triple-vaccinated HCWs infected during the B.1.1.529 (Omicron, black, *n* = 11) wave. (**C**) T cell responses against ancestral Wuhan Hu-1, B.1.617.2 (Delta), and B.1.1.529 (Omicron) S1 proteins plotted pair-wise for infection-naïve HCWs (blue, *n* = 10) or HCWs with laboratory-confirmed SARS-CoV-2 infection during the Wuhan Hu-1 (red, *n* = 3), B.1.1.529 (Omicron, black, *n* = 11), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, *n* = 6) waves. Statistical tests were performed using Prism 9.0. (A) Mann-Whitney *U* test, [(B) and (C)] Wilcoxon matched-pairs signed rank test.

cross-reactive S1 RBD binding antibodies against B.1.1.529 (Omicron) (Fig. 6C).

Hybrid immunity (the combination of prior infection and a single vaccine dose) significantly increased the S1 RBD binding antibodies against B.1.1.529 (Omicron) ($P < 0.0001$) compared with responses of infection-naïve HCWs, which were undetectable after a single vaccine dose. This increase was significantly greater for previously Wuhan Hu-1 infected than B.1.1.7 (Alpha) infected HCWs ($P < 0.0002$) (Fig. 6, B and C).

At 2 to 3 weeks after two vaccine doses, there was a leveling up of S1 RBD B.1.1.529 (Omicron) binding antibody, such that infection-naïve,

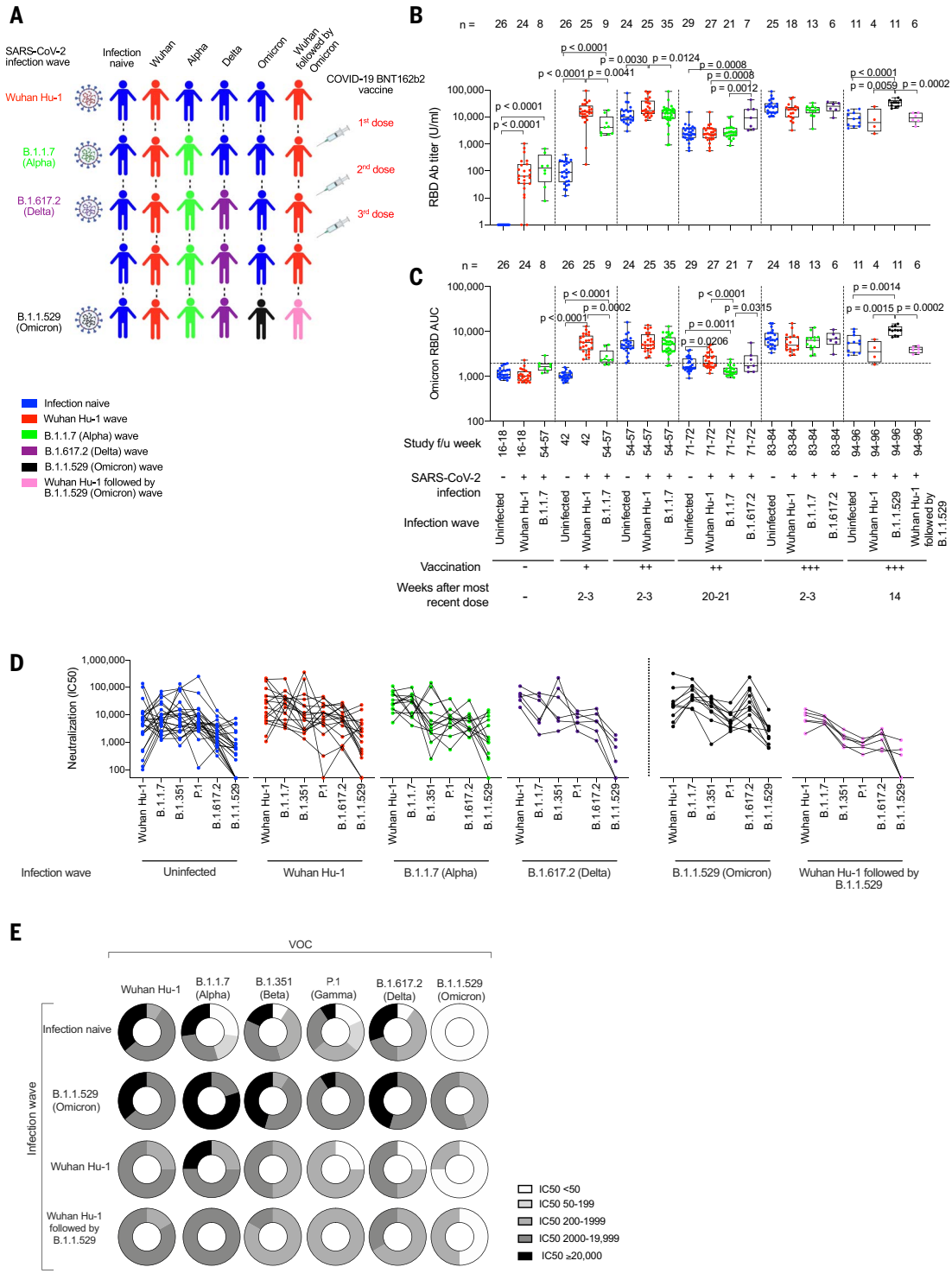


Fig. 6. SARS-CoV-2 infection imprints differential Ab cross-reactivity to VOCs. (**A**) Graphical summary depicting the SARS-CoV-2 infection and vaccination history of HCWs studied. Infection-naïve HCWs are indicated in blue. HCWs infected during the different waves are indicated as follows: ancestral Wuhan Hu-1 (red), B.1.1.7 (Alpha, green) and B.1.617.2 (Delta, purple), B.1.1.529 (Omicron, black), and Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink). (**B** and **C**) Serum Ab binding against ancestral Wuhan Hu-1 S1 RBD (B) and B.1.1.529 (Omicron) S1 RBD (C) in infection-naïve HCWs (blue, n = 11 to 29) or in HCWs with laboratory-confirmed SARS-CoV-2 infection during the ancestral Wuhan Hu-1 (red, n = 4 to 27), B.1.1.7 (Alpha, green, n = 8 to 35), B.1.617.2 (Delta, purple, n = 6 to 7), B.1.1.529 (Omicron, black, n = 11), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, n = 6) waves. Data are shown pre-vaccination and at time points after first, second, and third dose of vaccine. (**D**) Cross-reactive nAb IC_{50} against ancestral Wuhan Hu-1, B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), and B.1.1.529 (Omicron) live virus 2 to 3 weeks after third vaccine dose in infection-naïve HCWs (blue, n = 24) or HCWs with laboratory-confirmed SARS-CoV-2 infection during the Wuhan Hu-1 (red, n = 18), B.1.1.7 (Alpha, green, n = 13), or B.1.617.2 (Delta, purple, n = 6) waves and at 14 weeks (median, 14 weeks; IQR, 3 weeks) after third vaccine dose in HCWs with laboratory-confirmed SARS-CoV-2 during the B.1.1.529 (Omicron, black, n = 11) or Wuhan Hu-1 followed by B.1.1.529 (Omicron, pink, n = 6) waves. Data are plotted pair-wise for individual HCWs. (**E**) Doughnuts showing the relative proportion of HCWs with nAb IC_{50} of <50 (white), 50 to 199 (25% gray), 200 to 1999 (50% gray), 2000 to 19,999 (75% gray), and ≥20,000 (black) against ancestral Wuhan Hu-1, B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), and B.1.1.529 (Omicron) live virus in infection-naïve HCWs (n = 11) or HCWs with laboratory confirmed infection during the B.1.1.529 (Omicron, n = 11), Wuhan Hu-1 (n = 4), or Wuhan Hu-1 followed by B.1.1.529 (Omicron, n = 6) waves at 14 weeks (median, 14 weeks; IQR, 3 weeks) after the third vaccine dose. Statistical tests were performed using Prism 9.0. [(B) and (C)] Mann-Whitney U test.

previously Wuhan Hu-1 or B.1.1.7 (Alpha) infected HCWs made similar responses (Fig. 6, B and C).

However, 20 to 21 weeks after the second vaccine dose, differential B.1.1.529 (Omicron) RBD Ab waning was noted, with almost all (19/21) HCWs infected during the second B.1.1.7 (Alpha) wave no longer showing detectable cross-reactive antibody against B.1.1.529 (Omicron) RBD (Fig. 6C). This was distinct from HCWs infected by Wuhan Hu-1 during the first wave, who demonstrated a significantly higher cross-protective antibody response against B.1.1.529 (Omicron) RBD ($P < 0.0001$) (Fig. 6C). This indicates a profound differential impact of immune imprinting on B.1.1.529 (Omicron)–specific immune antibody waning between HCWs infected by Wuhan Hu-1 and B.1.1.7 (Alpha), as this differential is not seen in Ab responses to ancestral Wuhan Hu-1 spike S1 RBD (Fig. 6B).

Again, there was a leveling up back to similar B.1.1.529 (Omicron) RBD binding across infection-naïve and previously infected HCWs [Wuhan Hu-1, B.1.1.7 (Alpha), and B.1.617.2 (Delta)] 2 to 3 weeks after the third vaccine dose (Fig. 6, B and C). Fourteen weeks after the third vaccine dose, previously infection-naïve HCWs infected during the B.1.1.529 (Omicron) wave showed increased S1 RBD B.1.1.529 (Omicron) binding responses, but previously Wuhan Hu-1 infected HCWs did not, indicating that previously Wuhan Hu-1 infected individuals were immune imprinted to not boost antibody binding responses against B.1.1.529 (Omicron) despite having been infected by B.1.1.529 (Omicron) itself (Fig. 6C).

In fact, infection during the B.1.1.529 (Omicron) wave imprinted a consistent relative hierarchy of cross-neutralization immunity against VOCs across different individuals with potent cross-reactive nAb responses against B.1.1.7 (Alpha), B.1.351 (Beta), and B.1.617.2 (Delta) (Fig. 6, D and E). Comparative analysis of nAb potency for cross-neutralization of VOCs emphasized the impact of immune imprinting which effectively abrogates the nAb responses in those vaccinated HCWs infected during the first wave and then reinfected during the B.1.1.529 (Omicron) wave. The doughnuts in Fig. 6E highlight the extent to which the relative potency of nAb responses are attenuated in previously Wuhan Hu-1 infected HCWs.

Discussion

At this stage in the pandemic, there is a view that the global spread of B.1.1.529 (Omicron), through its association with a relatively milder disease phenotype and, possibly, a potential to boost vaccine immunity, may herald the transition into a new, endemic relationship (*28*). The case for vaccine-mediated immune preconditioning as key mediator of the attenuated phenotype is complex: Although functional neutralization by vaccine-primed sera is considerably blunted against B.1.1.529 (Omicron), three-dose vaccination efficacy against symptomatic disease holds up, in the 50 to 70% range (*6*–*8*). It has thus been proposed that immune protection may be supported by maintenance of relatively high T cell response frequencies to viral epitopes unperturbed by loss of antibody epitopes (*13*–*18*). A rationale for this T cell–mediated protection comes from animal studies showing the direct ability of SARS-CoV-2–specific T cells to curtail lung viral loads (*29*). This raised two key questions with respect to understanding and management of this wave: (i) Considering the very diverse patterns of antiviral immunity shown by ourselves and others to be determined by differential immune imprinting, how would differences in antigen exposure through infection and vaccination alter immune responses against B.1.1.529 (Omicron) at the level of binding antibody and nAb, MBC, and T cell responses? (ii) Is the immune response after infection during the B.1.1.529 (Omicron) wave primed and fully available to support protective immunity? We examined immunity to B.1.1.529 (Omicron) in a longitudinal HCW cohort, considering cross-reactive immunity primed by the varied spike exposures of three-dose vaccination with or without hybrid immunity from any of the Wuhan Hu-1, B.1.1.7 (Alpha), or B.1.617.2 (Delta) infection waves, and then the additive effect of actual infection during the B.1.1.529 (Omicron) wave. In the first part of this paper, we report patterns of response in differentially imprinted, triple-vaccinated HCWs. In the second part, we consider immune responses in those who went on to suffer infection during the B.1.1.529 (Omicron) wave despite triple-vaccination. There were several unexpected findings. Although it is known that cross-reactive antibody recognition is compromised by the mutations in B.1.1.529 (Omicron), it was surprising that this was so profoundly exacerbated by differential imprinting in those who were previously infected by either Wuhan Hu-1 or B.1.1.7 (Alpha). This adds an important dimension to global control of B.1.1.529 (Omicron) in light of the impact B.1.1.7 (Alpha) has had on the global pandemic—by May 2021, B.1.1.7 (Alpha) accounted for 67% of all cases across 149 countries (*30*). That previous SARS-CoV-2 infection history can imprint such a profound, negative impact on subsequent protective immunity is an unexpected consequence of COVID-19. While the notion that, generally, hybrid priming by infection and vaccination enhances immunity is widely agreed upon (*22*), imprinted patterns such as the specific combination of vaccination with infection during the first ancestral Wuhan Hu-1 wave followed by the B.1.1.529 (Omicron) wave require an additional term: "hybrid immune damping." Molecular characterization of the precise mechanism underpinning repertoire shaping from a combination of Wuhan Hu-1 or B.1.1.7 (Alpha) infection and triple-vaccination using ancestral Wuhan Hu-1 sequence, affecting immune responses to subsequent VOCs, will require detailed analysis of differential immune repertoires and their structural consequences. The impact of differential imprinting was seen just as profoundly in T cell recognition of B.1.1.529 (Omicron) S1, which was not recognized by T cells from any triple-vaccinated HCWs who were initially infected during the Wuhan Hu-1 wave and then reinfected during the B.1.1.529 (Omicron) wave. Notably, although B1.1.529 (Omicron) infection in triple-vaccinated previously uninfected individuals could indeed boost antibody, T cell, and MBC responses against other VOCs, responses to Omicron itself were reduced. This relatively poor immunogenicity against itself may help to explain why frequent B.1.1.529 (Omicron) reinfections with short time intervals between infections are proving a novel feature in this wave. It also concurs with observations that mRNA vaccination carrying the B.1.1.529 (Omicron) spike sequence (Omicron third-dose after ancestral sequence prime and boost) offers no protective advantage (*31*). Initial studies using acute serum samples after B.1.1.529 (Omicron) infection had indicated poor immunogenicity and a tendency to elicit only Omicron-specific responses in the unvaccinated and broader responses in those imprinted after COVID-19 vaccination (*32*, *33*), including unexpected patterns of combinations that appeared to ablate neutralizing responses to previously seen VOCs (*33*).

Our T cell analysis, which depended on processing of immunodominant epitopes from whole antigen, revealed a more profound deficit than others. Studies in which T cell responses of vaccinees against spike peptide megapools are screened show that, while there may be a 20% drop in response due to lost epitopes across the entire sequence, most of the response is maintained (*13*–*15*, *17*), albeit with a significant minority showing a completely ablated CD8 response to Omicron peptide pools (*17*). Other studies show that around a fifth of responders to peptide panels have a 50 to 70% drop in T cell response (*16*). Our approach was to evaluate T cell recognition using the dual approach of mapped epitope pools spanning the mutated regions and also whole, naturally processed antigen. We found the greatest impairment of T cell recognition when looking at epitope recognition after processing of whole antigen. Naturally processed epitopes from uptake of whole antigen would generally be considered more representative of the real-life situation and nearer to HLA-ligandome studies than synthetic megapools of several hundred overlapping peptides, which have the potential to drown out physiological response patterns

under the noise of responses from cryptic epitopes that may not feature in real-life natural responses. That is, megapool approaches can, by their nature, underestimate the extent of response ablation. The natural HLA-ligandome of peptides shown to be elicited by natural processing and HLAII presentation only partially overlaps epitopes mapped from overlapping synthetic peptide panels (*34*, *35*). Our immunization of mice with B.1.1.529 (Omicron) mutant epitopes confirmed that de novo T cell response repertoire can be elicited, but this is not necessarily the same as that generated during live infection.

The studies presented here have shown that the high global prevalence of B.1.1.529 (Omicron) infections and reinfections likely reflects considerable subversion of immune recognition at both the B and T cell, antibody binding, and nAb level, although with considerable differential modulation through immune imprinting. Some imprinted combinations, such as infection during the Wuhan Hu-1 and Omicron waves, confer particularly impaired responses.

Materials and methods

Study subjects

A total of 731 adult HCWs were recruited into the COVIDsortium bioresource in March 2020 (*19*–*24*) (fig. S1). A cross-sectional case-controlled substudy of 136 HCWs recruited 16 to 18 weeks after the March 2020 UK lockdown reported immunity to SARS-CoV-2 infection during the first UK wave (Wuhan Hu-1) (*19*). SARS-CoV-2 infection was determined by baseline, and weekly nasal RNA stabilizing swabs and Roche cobas SARS-CoV-2 reverse transcriptase polymerase chain reaction (RT-PCR) test and baseline and weekly Antibody testing for S1 using the IgG EUROIMMUN enzyme-linked immunosorbent assay (ELISA) and nucleocapsid using the ROCHE Elecsys electrochemiluminescence immunoassay (ECLIA). Antibody ratios >1.1 were deemed positive for the EUROIMMUN SARS-CoV-2 ELISA, and >1 was considered test positive for the ROCHE Elecsys anti-SARS-CoV-2 ECLIA, as evaluated by UK Health Security Agency (UKHSA), Porton Down, UK. A cross-sectional, case-controlled vaccine substudy cohort of 51 HCWs at a mean time point of 22 days (±2 days SD) after administration of the first dose of BNT162b2 vaccines reported immunity to vaccination in individuals with and without a history of prior SARS-CoV-2 infection during the Wuhan Hu-1 wave (*23*). The vaccine substudy recruited HCWs previously enrolled in the 16- to 18-week substudy. It included 25 HCWs (mean age: 44 years; 60% male) with previous lab-defined SARS-CoV-2 infection and 26 HCWs (mean age: 41 years; 54% male) with no laboratory evidence of SARS-CoV-2 infection throughout the initial 16-week longitudinal follow-up. HCWs were followed up longitudinally (n = 51) at a median time point of 20 days [7, interquartile range (IQR)] after administration of the second dose of BNT162b2 (*24*). An additional 358 HCWs were recruited at 55 to 57 weeks follow-up, 53 of whom were infected by the B.1.1.7 (Alpha) VOC during the second UK wave (*24*). At 71 to 72 weeks follow-up, 80 two-dose vaccinated HCWs were re-recruited who were either SARS-CoV-2 infection naïve (n = 27) or had been infected by Wuhan Hu-1 during the first wave (n = 31) or B.1.1.7 (Alpha) during the second UK wave (n = 22) (*24*). At 83 to 84 weeks, 62 HCWs had been recruited who were either SARS-CoV-2 infection–naïve (n = 25) or had been infected during the first Wuhan Hu-1 (n = 18), second B.1.1.7 (Alpha, n = 13), or third B.1.617.2 (Delta, n = 6) UK infection waves (table S1). All 62 had received a third dose of BNT162b2 at a median time point of 18 days (12, IQR) previously. Thirty-two HCWs were recruited at 94 to 96 weeks, median 14 weeks (3 weeks, IQR) after third-dose vaccination after the onset of the UK B.1.1.529 (Omicron) wave (table S6). This comprised 17 HCWs with PCR-confirmed infection during the Omicron wave. Eleven of these were previously infection-naïve, and six had prior Wuhan Hu-1 infection. A contemporaneous control group of HCWs not infected during the Omicron wave was also recruited; this comprised 11 infection-naïve HCWs and four with prior Wuhan Hu-1 infection. Lack of infection was confirmed by longitudinal N serology status (table S7).

Isolation of PBMCs

Peripheral blood mononuclear cells (PBMCs) were isolated from heparinized blood using Histopaque-1077 Hybri-Max^{T} (Sigma-Aldrich) density gradient centrifugation in SepMate tubes (Stemcell), as previously described (*20*, *23*, *24*). Isolated PBMCs were cryopreserved in fetal bovine serum (FBS) containing 10% dimethyl sulfoxide and stored in liquid nitrogen.

Isolation of serum

Whole blood samples in SST vacutainers (VACUETTE #455092) were clotted at room temperature for at least 1 hour and then centrifuged for 10 min at 800*g*. Serum was aliquoted and stored at −80°C for SARS-CoV-2 antibody detection.

Anti-N and anti-S1 serology

Anti-nucleocapsid and anti-spike antibody detection testing was conducted at UK Health Security Agency using the Roche cobas e801 analyzer. Anti-nucleocapsid antibodies were detected using the qualitative Roche Elecsys anti-SARS-CoV-2 ECLIA nucleocapsid assay (Roche ACOV2, product code: 09203079190) while anti-RBD antibodies were detected using the quantitative Roche Elecsys anti-SARS-CoV-2 ECLIA spike assay (Roche ACOV2S, product code: 09289275190). Assays were performed and calibrated as recommended by the manufacturer. Anti-N results are expressed as a cutoff index (COI) value based on the electrochemiluminescence signal of a two-point calibration, with results COI ≥ 1.0 classified as positive. Anti-spike results are expressed as units per milliliter (U/ml), similarly based on a two-point calibration and a reagent-specific master curve, with a quantitative range of 0.4 to 2500 U/ml. Samples with a value of ≥1.0 U/ml are interpreted as positive for spike antibodies, and samples exceeding >250 U/ml are automatically diluted by the analyzer.

Recombinant proteins

Wuhan Hu-1 SARS-CoV-2 S1 spike protein and B.1.617.2 (Delta) SARS-CoV-2 S1 spike protein (T19R, G142D, del156-157, R158G, L425R, T478K, D614G, P681R) (Z03485-1 and Z03612-1, respectively) were purchased from GenScript USA Inc. B.1.1.529 (Omicron) SARS-CoV-2 S1 spike (A67V, H69del, V70del, T95I, G142D, V143del, Y144del, Y145del, N211del, L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y) and RBD (G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H) proteins (REC32006 and REC32007, respectively) were purchased from the Native Antigen Company.

Peptides

Spike mapped epitope pool (MEP) comprises a previously described pool of eighteen 12- to 20mer peptide epitopes (*20*, *23*, *24*) (table S5A). B.1.1.529 (Omicron) and Wuhan Hu-1 peptide pools are comprised of peptides with the B.1.1.529 (Omicron) amino acid mutations and deletions and the respective Wuhan Hu-1 sequences (table S5B). They contain the predicted HLAII binding motifs, as determined by NetMHCIIpan4.0 (*34*) (table S9). Peptides were synthesized by GL Biochem Shanghai Ltd (China).

T cell assay by IFNγ ELISpot

Interferon-γ enzyme-linked immunosorbent spot (IFNγ ELISpot) assays were conducted as previously described (*20*, *23*, *24*). Precoated ELISpot plates (Mabtech 3420-2APT) were washed four times with phosphate-buffered saline (PBS), blocked for 1 hour (at room temperature) with supplemented RPMI1640 (GibcoBRL) [10% heat-inactivated FBS; 1% 100x penicillin, streptomycin and L-glutamine solutions (GibcoBRL)]. Two-hundred thousand PBMCs were seeded per well and stimulated for 18 to 22 hours at 37°C with SARS-CoV-2 recombinant protein [Wuhan Hu-1, B.1.617.2 (Delta), or B.1.1.529 (Omicron) SARS-CoV-2 S1 spike proteins (10 μg/ml)] or peptide pools

(10 μg per ml per peptide). Negative and positive plate controls were medium or anti-CD3 (Mabtech mAb CD3-2). ELISpot plates were developed with 1 μg/ml biotinylated anti-human IFNγ detection Ab conjugated to alkphosphatase (7-B6-1-ALP, Mabtech), diluted in PBS/0.5% FBS, adding 50 μl per well for 2 hours at room temperature followed by 50 μl per well BCIP/NBT-plus phosphatase substrate (Mabtech), 5 min (at room temperature). Plates were washed and dried before analysis on an AID classic ELISpot plate reader (Autoimmun Diagnostika GMBH, Germany). ELISpot data were analyzed in Microsoft Excel. The average of two culture media only wells was subtracted from all protein/peptide–stimulated wells, and any response that was lower in magnitude than 2 SD of the sample-specific control wells was not considered a stimulation-specific response. Results were expressed as difference in (delta) SFCs per 10^6 PBMCs between negative control and protein/peptide stimulation conditions. Results were excluded if negative control wells showed >100 SFCs per 10^6 PBMCs (n = 4) or cell viability was low, with <1000 SFCs per 10^6 PBMCs in anti-CD3 positive control wells (n = 5). Results were plotted using Prism 9.0 for Mac OS (GraphPad).

B cell ELISpots

Before B cell ELISpot assays, PBMCs were cultured for 5 days (37°C/5% CO_2) in 24-well plates, 500,000 cells per well containing 1 μg/ml TLR7/8 agonist R848 plus 10 ng/ml recombinant human IL-2 (Mabtech Human Memory B cell Stimpack 3660-1). After 4 days, PBMC stimulation ELISpot PVDF plates (Millipore MSIPS4W10) were coated with PBS, purified anti-human IgG MT91/145 (10 μg/ml, Mabtech 3850-3-250), and Wuhan Hu-1, B.1.617.2, or B.1.1.529 SARS-CoV-2 S1 spike proteins (10 μg/ml) and incubated at 4°C overnight. Plates were washed five times and blocked for 1 hour with RPMI1640 [supplemented with 10% heat-inactivated FBS, 1% 100x penicillin, streptomycin, and L-glutamine solutions (GibcoBRL)]. Pre-stimulated PBMCs were washed twice before seeding at 7500 to 15,000 cells per well for anti-human IgG coated wells and 15,000 to 150,000 cells per well for SARS-CoV-2 spike coated wells. Assays were run in duplicate. Plates were incubated at 37°C for 18 to 20 hours. For ELISpot development, plates were washed five times with PBS/0.05% Tween 20 (PBST) before incubation with 100 μl biotinylated anti-human IgG MT78/145 (Mabtech 3850-6-250), in PBS/0.5% FBS for 2 hours at room temperature. Plates were washed five times in PBST and incubated with 100 μl per well 1:1000 Streptavidin-ALP (Mabtech 3310-10-1000), in PBS/0.5% FBS for 1 hour at room temperature. Plates were then washed five times with PBST, and spots were developed by adding 100 μl per well BCIP/NBT substrate (Mabtech). Reactions were stopped by washing in tap water, and plates were dried before being analyzing on an AID classic ELISpot plate reader (Autoimmun Diagnostika GMBH, Germany). Analysis of ELISpot data was performed in Microsoft Excel. Spots counted for each well were adjusted for cell numbers seeded, and the average of PBS-only coated wells was subtracted from antigen-coated wells. The number of SARS-CoV-2 spike antigen–specific Ab secreting cells (ASCs) was expressed as a percent of the total number of IgG ASC.

B.1.1.529 (Omicron) RBD ELISA

Nunc 96-well immune ELISA plates were coated with 1 μg/ml of B.1.1.529 (Omicron) Spike RBD protein in carbonate buffer (Sigma Aldrich) for 2 hours at 37°C before washing with PBS (0.05% Tween) (PBST) and blocking at 37°C for 1 hour with PBS containing 1% bovine serum albumin (BSA). Plates were washed in PBST again before application of 50 μl of diluted sera to each well. All serum dilutions were run in duplicate, and a four-point dilution series was run for each sample. After overnight incubation at 4°C, plates were washed with PBST, and wells were incubated with 1:1000 dilution of Biotin Mouse Anti-human IgG (BD Pharmingen, 555785) at room temperature for 1 hour. Plates were washed again before application of 1:200 dilution of streptavidin horseradish peroxidase (HRP) (Bio-techne, DY998) for 30 min followed by a final wash and then assay development using 3,3′, 5,5;-tetramethylbenzidine (TMB) substrate (Sigma Aldrich, T0440). Color development was stopped after 5 min by the addition of 0.18 M H_2SO_4 and OD450nm values for each well measured using a FLUOstar Omega Plate Reader. Analysis of ELISA data was performed in Prism 9.0 for Mac OS (GraphPad). Data for serial dilutions were plotted, and area under the curve calculated for each individual serum sample.

Multiplex variant-specific IgG antibody measurement

Antibody titers against VOC-specific spike antigens (RBD or spike) were measured using the multiplex MesoScale Discovery (MSD) electrochemiluminescent immunoassay (V-Plex, MSD, Gaithersburg). IgG binding antibody to the RBD domain for the different VOCs were determined using the V-Plex Panel 22 (Catalogue number K15559U), which includes the RBD antigens of "wild-type" SARS-CoV-2, B.1.1.529/BA.1 (Omicron), B.1.1.7 (Alpha), B.1.351/B.1.351.1 (Beta), P.1 (Gamma), and AY.3/AY.4/B.1.617.2 (Delta). IgG binding antibody to the full spike protein of the different VOC were determined using V-Plex Panel 23 (Catalogue numb er K15567U) which includes spike antigens of "wild-type" SARS-CoV-2, AY.4.2 (Delta sublineage), B.1.1.529/BA.1 (Omicron), B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2/AY.3/ AY.5 (Delta and Delta sublineages), and AY.4 (Delta alternative sequence and sublineages) (*36*, *37*). A full list of each antigen and its included mutations can be found in the supplementary materials (table S2).

In brief, plates were run per manufacturer's instructions, with washing between incubations performed using a Biotek 405TS plate washer. Plates were blocked for 30 min with 5% BSA before the addition of samples diluted between 1:1000 and 1:100,000. Samples were incubated for 2 hours, followed by addition of the secondary anti-human IgG antibody (Sulfotag) for 1 hour. Read buffer was added to plates before reading immediately using the MSD QuickPlex SQ 120 platform. Only results from antigen spots within the detection range were used for the final analysis. Results were reported as arbitrary units per milliliter (AU/ml) determined against a seven-point calibration curve using serially diluted reference standard 1.

Authentic Wuhan Hu-1 SARS-CoV-2 and VOC variant titration

SARS-CoV-2 isolate stocks [including Wuhan Hu-1, B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), and B.1.1.529 (Omicron)] used in experiments (table S8) were prepared and titrated as previously described (*23*, *24*).

Authentic Wuhan Hu-1 SARS-CoV-2 and VOC microneutralization assays

SARS-CoV-2 microneutralization assays were carried out as described previously (*23*, *24*). VeroE6 cells were seeded in 96-well plates 24 hours before infection. Duplicate titrations of heat-inactivated participant sera were incubated with 3×10^4 focus-forming units of SARS-CoV-2 virus (TCID100) at 37°C for 1 hour. Serum/virus preparations were added to cells and incubated for 72 hours. Surviving cells were fixed in formaldehyde and stained with 0.1% (w/v) crystal violet solution [crystal violet was resolubilized in 1% (w/v) sodium dodecyl sulfate solution]. Absorbance readings were taken at 570 nm using a CLARIOStar Plate Reader (BMG Labtech). Negative controls of pooled pre-pandemic sera (collected before 2008) and pooled serum from neutralization-positive SARS-CoV-2 convalescent individuals were spaced across the plates. Absorbance for each well was standardized against technical positive (virus control) and negative (cells only) controls on each plate to determine percentage neutralization values. IC_{50} values were determined from neutralization curves. All authentic SARS-CoV-2 propagation and microneutralization assays were performed in a containment level 3 facility.

In silico epitope prediction for B.1.1.529 (Omicron) and BA.2

In silico predictions of HLA-DRB1 peptide-binding were performed using NetMHCIIpan-4.0

(*38*) on the basis of a peptide length of 15 amino acids (tables S9 and S13). HLA core binding sequences containing individual mutations were selected if contained within a peptide defined as a strong or weak binder by the NetMHCIIpan-4.0 default parameters of rank score <1% (threshold for strong binder) and rank score <5% (threshold for weak binder). For HLA-A, HLA-B, and HLA-C alleles, analysis was performed using NetMHCpan-4.1 on the basis of peptide lengths of 8, 9, and 10 amino acids (tables S10 to S12 and S14 to S16). Again, the default parameters of rank score ≤ 0.5% (threshold for strong binder) and ≤ 2% (threshold for weak binder) were used.

*HLA-DRB1*0401 transgenic T cell assays*

Studies using HLAII transgenics carrying DRB1*0401 in the context of a homozygous knockout for murine H2-Aβ have been previously described (*39*, *40*). Mice (7 to 8 weeks old) were immunized in one hind footpad with B.1.1.529 (Omicron) variant or Wuhan Hu-1 pools or peptides containing 10 μg of each peptide sequence in Hunters Titermax Gold adjuvant (Sigma Aldrich). Popliteal lymph nodes were collected at d10 and prepared as single-cell suspensions. IFNγ ELISpot assays were performed in triplicate in HL1 serum-free medium (Lonza) [supplemented with 1% 100x L-glutamine and 0.5% 100x penicillin/streptomycin solutions (GibcoBRL)]. PVDF ELISpot plates (Merck Millipore MSIPN4550) were coated with anti-mouse IFNγ capture antibody (Diaclone Murine IFN gamma ELISpot Set, 862.031.020) overnight before seeding 200,000 lymph node cells per well and stimulating (72 hours, 37°C with 5% CO_2) with peptide pools or individual SARS-CoV-2 Wuhan Hu-1 or variant peptides (10 μg per ml per peptide). Internal plate controls were culture media alone and staphylococcal enterotoxin B (SEB). Assays were developed using biotinylated anti-mouse IFNγ followed by streptavidin-alkaline phosphatase conjugate and BCIP/NTB substrate (Diaclone) before washing in tap water, drying, and analyzing using an AID classic ELISpot plate reader (Autoimmun Diagnostika GMBH, Germany). Analysis of ELISpot data was performed in Microsoft Excel. The average from three culture media wells was subtracted from peptide-stimulated wells, and any response that was <2 SD of the sample-specific control wells was not considered a peptide-specific response. Results were expressed as difference in (delta) SFCs per 10^6 PBMCs between the negative control and peptide stimulation conditions. Results were plotted using Prism 9.0 for Mac OS (GraphPad).

For transcriptomic analysis, lymph node cells were cultured with no peptide or with 10 μg/ml Wuhan Hu-1 or B.1.1.529 (Omicron) variant G142/del143-5 or Q493R/G496S/Q498R/N501Y/Y505H peptides. At 24 hours, cells were harvested and lysed for RNA extraction. RNA was extracted using an Agilent RNA microprep kit. cDNA was prepared using an RT2 first strand kit (Qiagen), and qPCR for target genes was performed using RT^2 profiler PCR array Mouse T Helper Cell Differentiation plates (Qiagen PAMM-503Z). Data were analyzed and plotted using the Qiagen GeneGlobe data analysis tool and genes up-regulated by peptide stimulation with $P > 0.05$ (by Student's t test) compared with no peptide stimulation were identified.

Statistics and reproducibility

Data were assumed to have a non-Gaussian distribution. Wilcoxon matched-pairs signed rank test and a Mann-Whitney U test were used for single, paired, and unpaired comparisons. Nonparametric tests were used throughout. $P < 0.05$ was considered significant. Prism 9.0 for Mac was used for analysis.

REFERENCES AND NOTES

1. S. S. A. Karim, Q. A. Karim, Omicron SARS-CoV-2 variant: A new chapter in the COVID-19 pandemic. *Lancet* **398**, 2126–2128 (2021). doi: 10.1016/S0140-6736(21)02758-6; pmid: 34871545
2. P. Elliott *et al*., Rapid increase in Omicron infections in England during December 2021: REACT-1 study. *Science* **375**, 1406–1411 (2022). doi: 10.1126/science.abn8347; pmid: 35133177
3. S. A. Madhi *et al*., Population immunity and Covid-19 severity with omicron variant in South Africa. *N. Engl. J. Med.* **386**, 1314–1326 (2022). doi: 10.1056/NEJMoa2119658; pmid: 35196424
4. K. P. Y. Hui *et al*., SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. *Nature* **603**, 715–720 (2022). doi: 10.1038/s41586-022-04479-6; pmid: 35104836
5. A. Sigal, Milder disease with Omicron: Is it the virus or the pre-existing immunity? *Nat. Rev. Immunol.* **22**, 69–71 (2022). doi: 10.1038/s41577-022-00678-4; pmid: 35046570
6. S. Collie, J. Champion, H. Moultrie, L.-G. Bekker, G. Gray, Effectiveness of BNT162b2 vaccine against omicron variant in South Africa. *N. Engl. J. Med.* **386**, 494–496 (2022). doi: 10.1056/NEJMc2119270; pmid: 34965358
7. H. F. Tseng *et al*., Effectiveness of mRNA-1273 against SARS-CoV-2 Omicron and Delta variants. *Nat. Med.* **28**, 1063–1071 (2022). doi: 10.1038/s41591-022-01753-y; pmid: 35189624
8. N. Andrews *et al*., Covid-19 vaccine effectiveness against the omicron (B.1.1.529) variant. *N. Engl. J. Med.* **386**, 1532–1546 (2022). doi: 10.1056/NEJMoa2119451; pmid: 35249272
9. D. Planas *et al*., Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. *Nature* **602**, 671–675 (2022). doi: 10.1038/s41586-021-04389-z; pmid: 35016199
10. L. Liu *et al*., Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* **602**, 676–681 (2022). doi: 10.1038/s41586-021-04388-0; pmid: 35016198
11. S. Cele *et al*., Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature* **602**, 654–656 (2022). doi: 10.1038/s41586-021-04387-1; pmid: 35016196
12. A. Muik *et al*., Neutralization of SARS-CoV-2 Omicron by BNT162b2 mRNA vaccine-elicited human sera. *Science* **375**, 678–680 (2022). doi: 10.1126/science.abn7591; pmid: 35040667
13. A. Tarke *et al*., SARS-CoV-2 vaccination induces immunological T cell memory able to cross-recognize variants from Alpha to Omicron. *Cell* **185**, 847–859.e11 (2022). doi: 10.1016/j.cell.2022.01.015; pmid: 35139340
14. Y. Gao *et al*., Ancestral SARS-CoV-2-specific T cells cross-recognize the Omicron variant. *Nat. Med.* **28**, 472–476 (2022). doi: 10.1038/s41591-022-01700-x; pmid: 35042228
15. J. Liu *et al*., Vaccines elicit highly conserved cellular immunity to SARS-CoV-2 Omicron. *Nature* **603**, 493–496 (2022). doi: 10.1038/s41586-022-04465-y; pmid: 35102312
16. V. Naranbhai *et al*., T cell reactivity to the SARS-CoV-2 Omicron variant is preserved in most but not all individuals. *Cell* **185**, 1041–1051.e6 (2022). pmid: 35202566
17. R. Keeton *et al*., T cell responses to SARS-CoV-2 spike cross-recognize Omicron. *Nature* **603**, 488–492 (2022). doi: 10.1038/s41586-022-04460-3; pmid: 35102311
18. R. R. Goel *et al*., Efficient recall of Omicron-reactive B cell memory after a third dose of SARS-CoV-2 mRNA vaccine. *Cell* **185**, 1875–1887.e8 (2022). doi: 10.1016/j.cell.2022.04.009; pmid: 35523182
19. T. A. Treibel *et al*., COVID-19: PCR screening of asymptomatic health-care workers at London hospital. *Lancet* **395**, 1608–1610 (2020). doi: 10.1016/S0140-6736(20)31100-4; pmid: 32401714
20. C. J. Reynolds *et al*., Discordant neutralizing antibody and T cell responses in asymptomatic and mild SARS-CoV-2 infection. *Sci. Immunol.* **5**, eabf3698 (2020). doi: 10.1126/sciimmunol.abf3698; pmid: 33361161
21. C. Manisty *et al*., Time series analysis and mechanistic modelling of heterogeneity and sero-reversion in antibody responses to mild SARS-CoV-2 infection. *EBioMedicine* **65**, 103259 (2021). doi: 10.1016/j.ebiom.2021.103259; pmid: 33662833
22. C. Manisty *et al*., Antibody response to first BNT162b2 dose in previously SARS-CoV-2-infected individuals. *Lancet* **397**, 1057–1058 (2021). doi: 10.1016/S0140-6736(21)00501-8; pmid: 33640038
23. C. J. Reynolds *et al*., Prior SARS-CoV-2 infection rescues B and T cell responses to variants after first vaccine dose. *Science* **372**, 1418–1423 (2021). doi: 10.1126/science.abh1282; pmid: 33931567
24. C. J. Reynolds *et al*., Heterologous infection and vaccination shapes immunity against SARS-CoV-2 variants. *Science* **375**, 183–192 (2022). doi: 10.1126/science.abm0811; pmid: 34855510
25. UK Office for National Statistics, Coronavirus (COVID-19) latest insights: Infections; https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/infections.
26. M. McCallum *et al*., N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332–2347.e16 (2021). doi: 10.1016/j.cell.2021.03.028; pmid: 33761326
27. G. Cerutti *et al*., Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* **29**, 819–833.e7 (2021). doi: 10.1016/j.chom.2021.03.005; pmid: 33789084
28. E. Callaway, Beyond Omicron: What's next for COVID's viral evolution. *Nature* **600**, 204–207 (2021). doi: 10.1038/d41586-021-03619-8; pmid: 34876665
29. B. Israelow *et al*., Adaptive immune determinants of viral clearance and protection in mouse models of SARS-CoV-2. *Sci. Immunol.* **6**, eabl4509 (2021). doi: 10.1126/sciimmunol.abl4509; pmid: 34623900
30. GISAID, Tracking of Variants; https://www.gisaid.org/hcov19-variants/.
31. M. Gagne *et al*., mRNA-1273 or mRNA-Omicron boost in vaccinated macaques elicits similar B cell expansion, neutralizing responses, and protection from Omicron. *Cell* **185**, 1556–1571.e18 (2022). doi: 10.1016/j.cell.2022.03.038; pmid: 35447072
32. S. I. Richardson *et al*., SARS-CoV-2 Omicron triggers cross-reactive neutralization and Fc effector functions in previously vaccinated, but not unvaccinated, individuals. *Cell Host Microbe* **30**, 880–886.e4 (2022). doi: 10.1016/j.chom.2022.03.029; pmid: 35436444
33. A. Rössler, L. Knabl, D. von Laer, J. Kimpel, Neutralization profile after recovery from SARS-CoV-2 omicron infection. *N. Engl. J. Med.* **386**, 1764–1766 (2022). doi: 10.1056/NEJMc2201607; pmid: 35320661
34. M. D. Knierman *et al*., The human leukocyte antigen class II immunopeptidome of the SARS-CoV-2 spike glycoprotein. *Cell Rep.* **33**, 108454 (2020). doi: 10.1016/j.celrep.2020.108454; pmid: 33220791
35. R. Parker *et al*., Mapping the SARS-CoV-2 spike glycoprotein-derived peptidome presented by HLA class II on dendritic cells. *Cell Rep.* **35**, 109179 (2021). doi: 10.1016/j.celrep.2021.109179; pmid: 34004174
36. Meso Scale Diagnostics, V-PLEX SARS-CoV-2 Panel 22 (IgG) Kit; https://www.mesoscale.com/products/v-plex-sars-cov-2-panel-22-igg-kit-k15559u/.
37. Meso Scale Diagnostics, V-PLEX SARS-CoV-2 Panel 23 (IgG) Kit; https://www.mesoscale.com/products/v-plex-sars-cov-2-panel-23-igg-kit-5-pl-k15567u/.
38. B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of

MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020). doi: 10.1093/nar/gkaa379; pmid: 32406916
39. K. J. Quigley *et al*., Chronic infection by mucoid *Pseudomonas aeruginosa* associated with dysregulation in T-cell immunity to outer membrane porin F. *Am. J. Respir. Crit. Care Med.* **191**, 1250–1264 (2015). doi: 10.1164/rccm.201411-1995OC; pmid: 25789411
40. C. Reynolds *et al*., T cell immunity to the alkyl hydroperoxide reductase of *Burkholderia pseudomallei*: A correlate of disease outcome in acute melioidosis. *J. Immunol.* **194**, 4814–4824 (2015). doi: 10.4049/jimmunol.1402862; pmid: 25862821

ACKNOWLEDGMENTS

The authors thank HCW participants for participating in the study and the research teams involved in recruitment, obtaining consent, and sampling the HCW participants. The SARS-CoV-2 Wuhan Hu-1 Human 2019-nCoV Isolate EVA catalog code 026V-03883 was obtained from European Virus Archive Global (EVAg), Charité – Universitätsmedizin Berlin. The SARS-CoV-2 B.1.1.7 isolate was obtained from the National Institute for Biological Standards and Control (NIBSC), thanks to the contribution of PHE Porton Down and S. Funnell. The nCoV19 isolate/UK ex South African/2021 lineage B.1.351 EVA catalog code 04V-04071 was obtained from EVAg, PHE Porton Down. P.1, B.1.617.2, and B.1.1.529. isolates were purchased from EVAg. Figures 1A, 4A, and 6A in this manuscript were created with BioRender.com. **Funding:** R.J.B. and D.M.A. are supported by MRC (MR/S019553/1, MR/R02622X/1, MR/V036939/1, and MR/W020610/1), NIHR Imperial Biomedical Research Centre (BRC):ITMAT, Cystic Fibrosis Trust SRC (2019SRC015), NIHR EME Fast Track (NIHR134607), NIHR Long Covid (COV-LT2-0027), Innovate UK (SBRI 10008614), and Horizon 2020 Marie Skłodowska-Curie Innovative Training Network (ITN) European Training Network (860325). Á.M. is supported by MRC (MR/W020610/1), NIHR EME Fast Track (NIHR134607), Rosetrees Trust, the John Black Charitable Foundation, and Medical College of St Bartholomew's Hospital Trust. The COVIDsortium is supported by funding donated by individuals, charitable trusts, and corporations including Goldman Sachs, Kenneth C Griffin, the Guy Foundation, GW Pharmaceuticals, Kusuma Trust, and Jagclif Charitable Trust and enabled by Barts Charity with support from UCLH Charity. Wider support is acknowledged on the COVIDsortium website (https://covid-consortium.com/). Institutional support from Barts Health NHS Trust and Royal Free NHS Foundation Trust facilitated study processes, in partnership with University College London and Queen Mary University of London. J.C.M., C.M., and T.A.T. are directly and indirectly supported by the University College London Hospitals (UCLH) and Barts NIHR Biomedical Research Centres and through the British Heart Foundation (BHF) Accelerator Award (AA/18/6/34223). T.A.T. is funded by a BHF Intermediate Research Fellowship (FS/19/35/34374). The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. **Ethics statement:** The COVIDsortium HCW bioresource is registered on ClinicalTrials.gov (NCT04318314) and approved by the UK National Research Ethics Service (20/SC/0149). Subjects gave written, informed consent, and the study conformed to the Helsinki Declaration principles. Mouse experiments were performed under UK Home Office Legislation and the Animals (Scientific Procedures) Act 1986 and Project License P809B6A94 granted for this work. **Author contributions:** R.J.B. conceptualized, designed, and led the study reported here. R.J.B. and D.M.A. designed and supervised T cell and B cell experiments and HLAII transgenic experiments. Á.M. designed and supervised the nAb experiments. T.B. and A.S. supervised whole spike and S1 RBD IgG and N IgG/IgM studies. C.J.R. developed, performed, and analyzed T cell and memory B cell experiments and HLAII transgenic and transcriptomic experiments. J.M.G. and C.P. developed, performed, and analyzed nAb experiments. A.D.O. performed and A.Se. analyzed whole spike, RBD, and N antibody assays. T.B., C.M., Á.M., T.A.T., J.C.M., and M.N. conceptualized and established the COVIDsortium HCW cohort. R.J.B, T.A.T., J.C.M., and C.M. conceptualized and designed the vaccine substudy cohort. G.J., N.F., C.M., T.A.T., and J.C.M. recruited HCWs and collected samples. C.J.R., D.M.S., K.-M.L., F.P.P., S.L., and D.K.B. processed HCW samples. C.J.R., J.M.G., C.P., Á.M., A.D.O., D.M.A., and R.J.B. analyzed the data. C.J.R., J.M.G., C.P., A.D.O., C.M., T.A.T., J.C.M., A.S., T.B., Á.M., D.M.A., and R.J.B. interpreted the data. R.J.B. and D.M.A. wrote the original and revised manuscripts with input from the authors. All authors reviewed and approved the manuscript and figures. **Competing interests:** R.J.B. and D.M.A. are members of the Global T cell Expert Consortium and have consulted for Oxford Immunotec outside the submitted work. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. The SARS-CoV-2 Wuhan Hu-1 Human 2019-nCoV, B.1.351, P.1, B.1.617.2, and B.1.1.529 isolates were obtained under material agreements with EVAg, France. The SARS-CoV-2 B.1.1.7 isolate was obtained under a material agreement with NIBSC, UK.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq1841
Figs. S1 to S6
Tables S1 to S17
COVIDsortium Investigators
COVIDsortium Immune Correlates Network
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 23 March 2022; accepted 6 June 2022
10.1126/science.abq1841

RESEARCH ARTICLE

CANCER IMMUNOLOGY

TCR-Vγδ usage distinguishes protumor from antitumor intestinal γδ T cell subsets

Bernardo S. Reis[1]*, Patrick W. Darcy[1], Iasha Z. Khan[1], Christine S. Moon[2], Adam E. Kornberg[2], Vanessa S. Schneider[1,3], Yelina Alvarez[1], Olawale Eleso[1], Caixia Zhu[1]†, Marina Schernthanner[1], Ainsley Lockhart[1], Aubrey Reed[1], Juliana Bortolatto[1], Tiago B. R. Castro[1], Angelina M. Bilate[1], Sergei Grivennikov[4], Arnold S. Han[2], Daniel Mucida[1,5]*

γδ T cells represent a substantial fraction of intestinal lymphocytes at homeostasis, but they also constitute a major lymphocyte population infiltrating colorectal cancers (CRCs); however, their temporal contribution to CRC development or progression remains unclear. Using human CRC samples and murine CRC models, we found that most γδ T cells in premalignant or nontumor colons exhibit cytotoxic markers, whereas tumor-infiltrating γδ T cells express a protumorigenic profile. These contrasting T cell profiles were associated with distinct T cell receptor (TCR)–Vγδ gene usage in both humans and mice. Longitudinal intersectional genetics and antibody-dependent strategies targeting murine γδ T cells enriched in the epithelium at steady state led to heightened tumor development, whereas targeting γδ subsets that accumulate during CRC resulted in reduced tumor growth. Our results uncover temporal pro- and antitumor roles for γδ T cell subsets.

Intestinal intraepithelial lymphocytes (IELs) make up a large T cell population located at the critical interface between the core of the body and the intestinal lumen, which is constantly exposed to food, commensal microbes, and pathogens. Previous observations have suggested an important role for IELs as a first line of immunity against pathogens (*1*–*4*). Among the main IEL subsets in mice or humans are T cells harboring the γδ T cell receptor (TCR). These γδ IELs are finely tuned to local epithelial signals and perform epithelial surveillance, which is modulated through cross-talk with the intestinal epithelium (*3*, *5*). In addition to their role in immune surveillance against enteric infections, γδ T cells have been associated with antitumor activity, including in human colorectal cancer (CRC) (*6*–*11*). CRC is currently the second most deadly cancer in the United States (according to the American Cancer Society, Inc.). The cumulative risk of inflammatory bowel disease (IBD) patients developing CRC can reach 20%; however, most CRC cases develop in patients without underlying inflammation. In both scenarios, tumor-elicited inflammation triggered by epithelial disturbances and microbial invasion is essential for the survival of malignant cells and tumor growth (*12*–*15*). In this study, we addressed whether epithelial-resident γδ T cell subsets could prevent CRC development and whether phenotypically and functionally distinct γδ T cell subsets play a contrasting role—accelerating tumor progression.

[1]Laboratory of Mucosal Immunology, The Rockefeller University, New York, NY 10065, USA. [2]Department of Medicine, Division of Digestive and Liver Diseases, Columbia University, New York, NY 10032, USA. [3]Department of Biochemistry and Molecular Biology, Federal University of Parana, Curitiba, PR, Brazil. [4]Department of Medicine and Department of Biomedical Sciences, Cedars-Sinai Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA. [5]Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA.
***Corresponding author. Email: breis@rockefeller.edu (B.S.R.); mucida@rockefeller.edu (D.M.)**
†Present address: Key Laboratory of Medical Molecular Virology, School of Basic Medical Sciences, Shanghai Medical College, Fudan University, Shanghai 200032, China.

Distinct profiles of infiltrating γδ T cells in both human and murine CRC

Interleukin-17 (IL-17)–producing γδ T cells preferentially use oxidative-phosphorylation metabolism and have been associated with increased tumor burden and poor survival in human cancers (*16*–*18*). By contrast, glycolytic interferon-γ (IFN-γ)–producing γδ T cells have been associated with protection against tumors and a better prognosis (*19*, *20*). To characterize the gene signature of γδ T cells found in the intestines of CRC patients, we sorted γδ T cells from surgically dissected tumors and adjacent nontumor areas, stimulated in vitro, and performed single-cell RNA sequencing (scRNA-seq) using the 10X genomics platform. We collected 1825 tumor-infiltrating and adjacent cells from five patients displaying at least 100 viable γδ T cells per region (out of seven patients screened) (table S1 and fig. S1, A and B). Pooled analysis of 716 cells from tumors and 1109 from nontumor areas revealed 10 different clusters (Fig. 1A). Tumor-infiltrating γδ T cells showed an overall increased cytokine signature (gene ontology: 0005126), including IL-17–producing γδ T cell–related genes, such as CD9 and LGALS3 (*21*) enriched in clusters 2 and 5, when compared with cells isolated from adjacent nontumor areas (Fig. 1, B and C). By contrast, γδ T cells isolated from adjacent nontumor areas presented a cytotoxic-related profile, including the expression of GZMB and CXCR3, enriched in clusters 0 and 1, as well as the glycolysis-associated genes ENO1 and ALDOA (Fig. 1, B and C, and fig. S1, B and C), an overall profile resembling IFN-γ–producing γδ T cells (*8*). Clonal analysis of γδ T cells indicated clonal expansion in both nontumor and tumor areas, with an enrichment for Vδ1 gene usage by tumor-infiltrating γδ T cells, whereas cells isolated from both areas displayed preference for Vγ4 (Fig. 1, D and E, and fig. S1, D to F). Additionally, we found a correlation between gene signature clusters and clonal expansion, which confirms that expanded γδ T cell clones from tumors are enriched for IL-17–producing γδ T cell–related signature (clusters 2 and 5), whereas expanded clones in adjacent areas were enriched for IFN-γ–producing γδ T cell signatures (clusters 0 and 1) (Fig. 1E). Moreover, expanded clones related to the major gene expression clusters found in nontumor areas (TRVG4_TRDV1 bias in cluster 1 and TRGV8_TRDV3 in cluster 0) were reduced in tumor sites, whereas clones related to the major gene expression clusters found in tumor areas (TRVG4_TRDV1 bias in clusters 2 and 5 and TRVG9_TRDV1 in cluster 3) were reduced in nontumor areas (Fig. 1E). The above analyses indicate that γδ T cells enriched in human CRC areas share similarities with protumorigenic IL-17–producing γδ T cells, whereas cells found in tumor-adjacent areas display a cytotoxic T lymphocyte (CTL)– or IFN-γ–producing γδ T cell signature, overall related to antitumor function.

To functionally and mechanistically assess the role of γδ T cells in CRC, we used two distinct CRC mouse models: the chemical azoxymethane–dextran sodium sulphate (AOM-DSS) colitis-associated CRC (CAC) (*22*) and the genetically inducible adenomatosis polyposis coli (APC) deficiency (*14*, *23*, *24*) (fig. S2A). Our initial characterization of γδ T cells in naïve animals confirmed a dominance by Vγ1- and Vγ7-expressing subsets, followed by $Vγ4^+$ and $Vγ6^+$ subsets at a low frequency (*25*–*27*), and changes in the expression of CD8αα homodimers according to proximal-distal and villus-crypt axes (fig. S2, B to D). In CAC (AOM-DSS), we observed that tumor-infiltrating γδ T cells adopt both potentially anti- and protumor phenotypes, with increased expression of CD107α and IFN-γ [associated with antitumor responses (*13*, *28*)] as well as IL-17 [associated with protumorigenic function (*14*, *28*)] (fig. S2E). $CD4^+$ $TCRαβ^+$ cells also display IFN-γ– and IL-17–producing phenotypes, but in contrast to γδ T cells, expression levels upon restimulation are similar between tumor and nontumor tissue sites (fig. S2F). In the CAC model, ~40% of IL-17–producing cells within the tumor are γδ T cells

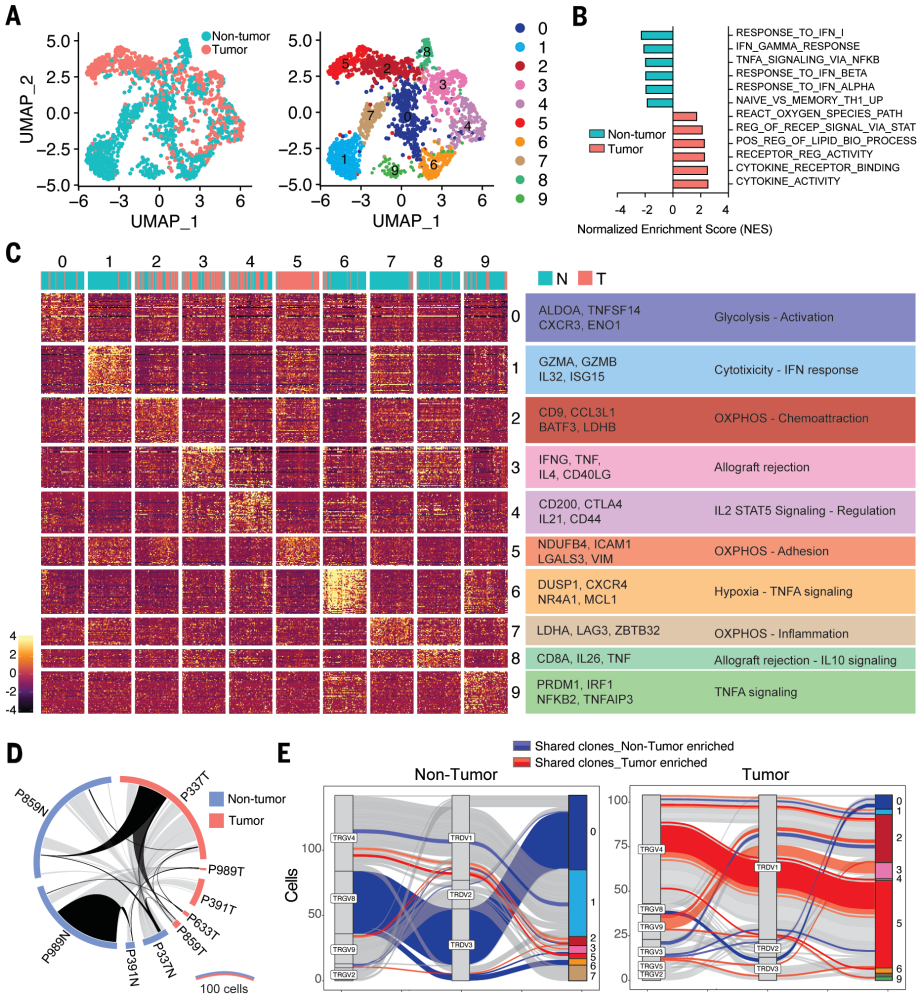


Fig. 1. Profiling of human γδ T cells in patients with CRC identifies tissue-specific subsets. γδ$^+$ T cells were sorted from tumor and adjacent (nontumor) areas of human CRC colonic resection tissue and processed for 10X genomics RNA and TCR sequencing. Cells were stimulated with phorbol myristate acetate (PMA) and ionomycin before RNA-seq. (**A**) Uniform manifold approximation and projection (UMAP) plot colored by tissue (left) and gene expression cluster (right) of γδ$^+$ T cells. (**B**) Gene set enrichment analysis (GSEA) of γδ$^+$ T cells recovered from nontumor (blue) and tumor (red) areas. (**C**) Gene expression heatmap and characterization of γδ$^+$ clusters based on GSEA hallmarks. Contribution of nontumor (blue) and tumor (red) cells in gene expression clusters is depicted above the heatmap. (**D**) Circos plot of shared clones between tissues (light gray) and between patients (black), based on amino acid CDR3 sequence. (**E**) Parallel plots depicting V gene usage and gene expression clusters of expanded clones found in nontumor (left) and tumor (right) areas. Clones (represented by lines) shared between tissues are colored.

and ~30% are CD4$^+$ T cells [T helper 17 (T_H17) cells], whereas in nontumor areas, they correspond to ~35 and 20%, respectively (fig. S2G). Similar to what was observed in naïve mice, γδ T cells in healthy colon tissue display high expression of CD8αα, whereas a CD8α$^-$ population expressing the exhaustion marker PD-1$^+$ is prominent among tumor-infiltrating γδ T cells (fig. S2, H and I). PD-1 expression segregated cells with potential antitumor (PD-1$^-$: IFN-γ$^+$) or protumor phenotypes (PD-1$^+$: IL-17$^+$ and IFN-γ$^-$) (fig. S2, J and K).

To gain information about γδ T cell dynamics during CRC progression, we longitudinally analyzed CRC mice harboring inducible APC-floxed alleles, *Cdx2*$^{Cre\text{-}ER}$ x APC$^{fl/fl}$ (i*Cdx2*$^{\Delta APC}$). Although virtually absent in naïve mice, we found that γδ T cells expressing IL-17 and PD-1 accumulate in the colonic tissue of i*Cdx2*$^{\Delta APC}$ mice upon tamoxifen injection, whereas the frequency of CD8α$^+$ γδ T cells decreases (Fig. 2A). Similar to our observations in the CAC model, the frequency of γδ T cells producing IFN-γ and IL-17 also increases among tumor-infiltrating lymphocytes (fig. S2L). Additionally, tumor-infiltrating PD-1$^+$ (CD8α$^-$) preferentially express IL-17, whereas PD-1$^-$ (CD8α$^+$) preferentially secrete IFN-γ, with no overlap between IL-17– and IFN-γ–producing γδ cells. (Fig. 2, B to D, and fig. S2, M and N). CD4$^+$ T cells did not show significant differences in IL-17 and IFN-γ production between tumor or adjacent nontumor areas (fig. S2O). In the APC loss model, ~60% of IL-17–producing cells within the tumor are γδ T cells, whereas T_H17 cells constitute only 20% (fig. S2P). Outside tumor areas, almost 70% of IL-17–producing cells are γδ T cells and ~10% are T_H17 cells upon in vitro restimulation (fig. S2P). To further characterize tumor-infiltrating γδ T cells in the CAC model, we sorted these cells on the basis of PD-1 expression and performed RNA-seq analysis. Analogous to our observations in humans, murine CD8α$^+$PD-1$^-$ γδ T cells (blue) displayed increased expression of cytotoxic-related genes (e.g., *Gzmb*, *Gzma*, *Lgals3*, and *Lag3*), whereas CD8α$^-$PD-1$^+$ γδ T cells (red) showed increased expression of genes associated with IL-17 and protumorigenic responses (*15*), such as *Il-17*, *Rorc*, and *Il1r1* (Fig. 2E). These results suggest opposing phenotypes among γδ T cells found at steady state or in nontumor areas, versus γδ T cells that accumulate during CRC in mice and humans.

Our bulk RNA-seq analysis suggested divergent TCR usage between tumor-infiltrating CD8α$^-$PD-1$^+$ cells enriched in *TcrgV6* transcripts and CD8α$^+$PD-1$^-$ γδ T cells enriched in *TcrgV7* transcripts (Fig. 2E). To directly investigate the TCR repertoire of murine tumor-infiltrating γδ T cells, we single-cell sorted γδ T cells from the tumor and adjacent nontumor areas of CRC mice subjected to either AOM-DSS or APC loss and performed single-cell TCR sequencing analysis (scTCR-seq). In the APC loss model, γδ T cells from nontumor areas are clonally diverse, whereas tumor-infiltrating γδ T cells show noticeable clonal expansions (Fig. 2F and data S1). In the AOM-DSS model, we observed some clonal expansion within the Vγ7 subset found in nontumor areas (Fig. 2F and fig. S2Q), an effect possibly related to the chronic inflammatory nature of this model (*29*), although most clonal expansions were also found in the tumor areas (Fig. 2F and data S2). The expanded clones found in tumor areas in both models were primarily Vγ6$^+$Vδ1$^+$clones (*Trgv6* and *Trdv4* rearrangements), which are

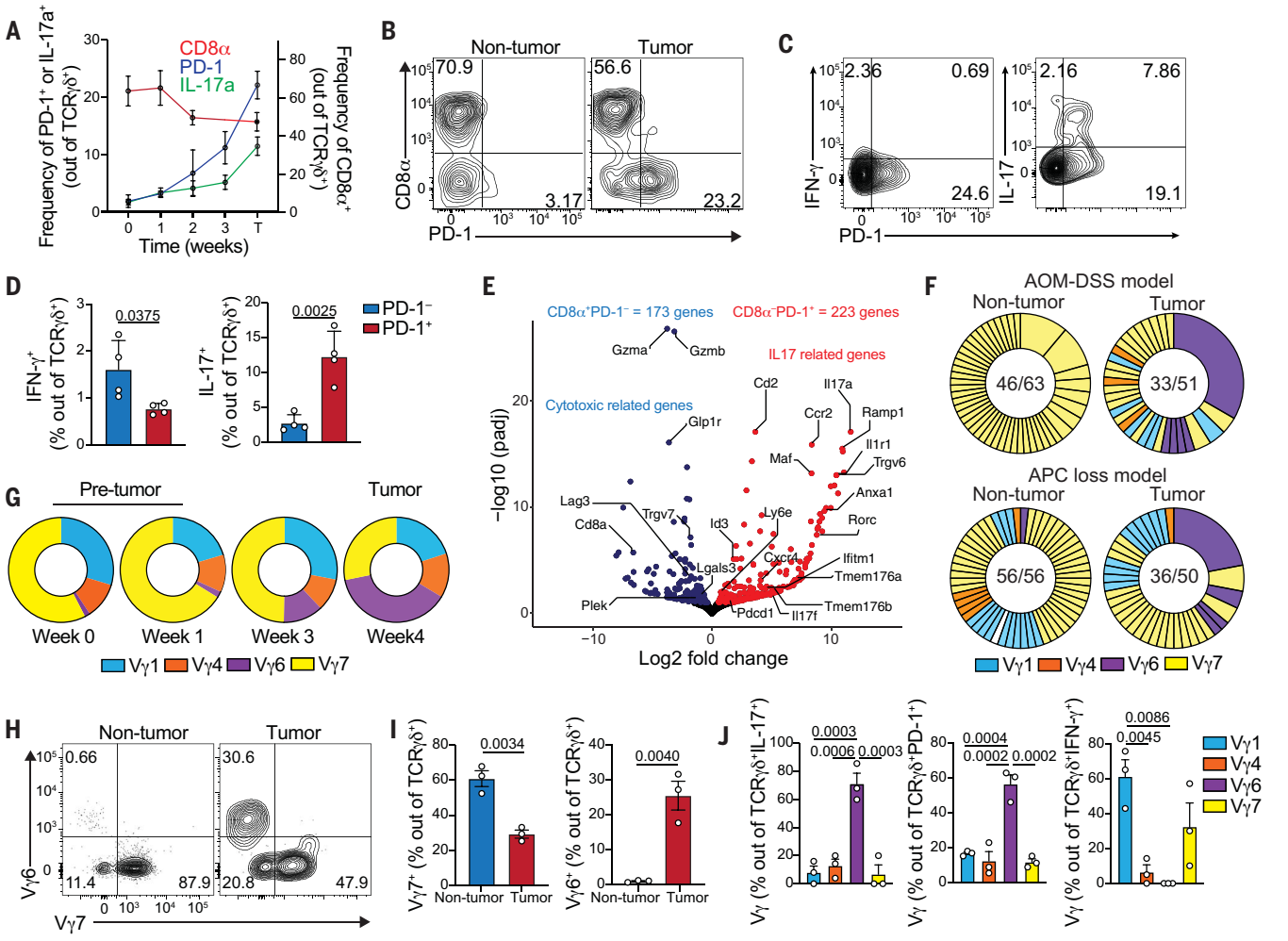


Fig. 2. Profiling tumor-infiltrating γδ T cells in CRC models reveals distinct subsets. *iCdx2*$^{\Delta APC}$ animals were treated with tamoxifen and euthanized at indicated times for analysis of large intestine IELs (weeks 0 to 3) and tumor areas (week 4 or T). (**A**) Frequency of CD8α$^+$ (right axis) and PD-1$^+$ or IL-17$^+$ (left axis) cells among TCRγδ$^+$ T cells from the large intestine before (week 0) and after (weeks 1, 2, 3, and T) tamoxifen treatment. (**B**) Representative dot plot of CD8α$^+$ and PD-1$^+$ among TCR γδ$^+$ cells at 4 weeks after tamoxifen administration. (**C** and **D**) Representative dot plot (C) and frequency (D) of IFN- γ$^+$ (left) or IL-17$^+$ (right) among tumor-infiltrating PD-1$^+$ or PD-1$^-$ TCR γδ$^+$ cells (APC loss model). (**E**) Volcano plot of differentially expressed genes from RNA-seq analysis of sorted CD8α$^+$PD-1$^-$ (blue) or CD8α$^-$PD-1$^+$ (red) TCR γδ$^+$ cells isolated from the tumors of mice subjected to the AOM-DSS model. (**F**) scTCR-seq of γδ T cells from tumor or nontumor colonic tissues of four mice subjected to the AOM-DSS (top) and four mice from the APC loss (bottom) models. Numbers in the centers of the pie charts represent the number of clones (based on CDR3 aa sequence) per total cells sequenced. Expanded clones are fused. Clones are colored based on Vγ usage. Purple clones represent expanded Vγ6Vδ1 cells. (**G**) Pie charts of Vγ frequency among TCRγδ$^+$ cells from large intestine tissue before (week 0) and after (weeks 1, 3, and 4) tamoxifen treatment (APC loss model). (**H** and **I**) Vγ usage by γδ T cells from tumor or nontumor colonic tissue of mice subjected to the AOM-DSS protocol. Representative dot plot (H) and frequency (I) of Vγ6$^+$ and Vγ7$^+$ among TCRγδ$^+$ cells. (**J**) Frequency of Vγ1$^+$, Vγ4$^+$, Vγ6$^+$, and Vγ7$^+$ among tumor-infiltrating TCRγδ$^+$ cells expressing IL-17, PD-1, or IFN-γ. Representative data are from two experiments with three to four animals per group. RNA-seq and TCR-seq data are from pooled tumors. For cytokine staining, cells were stimulated with PMA and ionomycin. Statistical *P* value differences are indicated. Statistics are by two-tailed *t* test [(E) and (I)] and one-way analysis of variance (ANOVA) with Dunnett's multiple comparison test (J). Error bars indicate SEMs.

rarely observed in nontumor areas or in the early phase of tumor development (Fig. 2, F and G, and fig. S2R). Flow cytometry analysis confirmed the relative increase of Vγ6$^+$ cells at the expense of decreased Vγ7$^+$ γδ T cells in the tumor areas (Fig. 2, H and I). As suggested by the bulk RNA-seq analysis, tumor-infiltrating Vγ7$^+$ and Vγ1$^+$ T cells isolated from mice subjected to AOM-DSS are mostly PD-1$^-$ and IFN-γ$^+$, whereas Vγ6$^+$ cells express PD-1 and IL-17 (Fig. 2J). Overall, tumor-infiltrating PD-1$^+$ and IL-17$^+$ γδ T cells are composed of Vγ6 (56.5 ± 5.2% and 71.2 ± 7.5%, respectively), whereas IFN-γ$^+$ γδ T cells are composed of Vγ1 (61.1 ± 9.6%) and Vγ7 (32.4 ± 13.6%) in the CAC model (Fig. 2J). Similar Vγ distribution was observed in tumor-infiltrating TCRγδ$^+$ T cells isolated from CRC mice subjected to APC loss (fig. S2S). Immunofluorescence imaging of the colonic tissue from tamoxifen-treated *iCdx2*$^{\Delta APC}$ mice confirmed the preferential accumulation of PD-1–expressing γδ T cells in tumor areas (fig. S2T). The analyses above suggest that tumor-infiltrating murine γδ T cells are composed of two major functional groups based on Vγ usage: polyclonal Vγ7$^+$ and Vγ1$^+$ cells exhibiting an antitumor cytotoxic program (resembling IFN-γ–expressing human subsets enriched in clusters 0 and 1; Fig. 1A) and clonally expanded Vγ6$^+$Vδ1$^+$ cells that express PD-1 and secrete IL-17 (resembling IL-17–expressing human subsets enriched in clusters 2 and 5; Fig. 1A).

Antitumor activity by epithelium-resident murine Vγ1$^+$ and Vγ7$^+$ γδ T cells

To address whether the phenotype of epithelium-resident TCRγδ T cells plays a role in restricting CRC development, we first subjected *Trdc*$^{-/-}$ knockout mice, which are deficient in γδ T cells, to the AOM-DSS model (fig. S3, D to F). Because *Trdc*$^{-/-}$ animals are more susceptible to the experimental DSS regimen (*30*), we used a lower DSS concentration (1%), which does not result in noticeable inflammation or tumor development in wild-type control mice (fig. S3, E and A to C). In line with previous literature (*10*), after the 1% DSS regimen, *Trdc*$^{-/-}$ mice displayed enhanced inflammation and significant increase in tumor development, again suggesting an antitumor or anti-inflammatory role for intestinal epithelium-resident γδ T cell populations (fig. S3, E and F). Because total *Trdc*-deficiency also prevents the accumulation of γδ cells during tumor progression, we next aimed to preferentially restrict epithelium-resident γδ T cell function by inducible hemizygous or homozygous inactivation of *Scl2a1* (i*Trdc*$^{Scl2a1fl/+}$ or i*Trdc*$^{\Delta Scl2a1}$) encoding the glucose transporter Glut1. We have previously reported that epithelium-resident γδ T cells control early invasion by *Salmonella* Typhimurium through a metabolic switch toward glycolysis that is dependent on Glut1 expression (*3*), an observation related to recent findings showing that IFN-γ–secreting, but not IL-17–secreting, γδ T cells are glycolytic and exert antitumor activity (*31*). Our analysis of human samples described above also pointed to glycolytic γδ T cells that were preferentially found in the tumor-adjacent areas. We then subjected i*Trdc*$^{\Delta Scl2a1}$, i*Trdc*$^{Scl2a1fl/+}$, and littermate control mice to the AOM-DSS model (Fig. 3A). Early targeting of Glut1 in γδ T cells results in higher tumor number and load when compared with that in control animals, without affecting tumor size (Fig. 3B). Moreover, although Glut1 inactivation does not lead to changes in CD8α, PD-1, or IL-17 expression by γδ T cells, we detected a significant reduction in IFN-γ production by tumor-infiltrating γδ T cells in both i*Trdc*$^{Scl2a1fl/+}$ and i*Trdc*$^{\Delta Scl2a1}$ mice when compared with littermate controls (Fig. 3, C and D, and fig. S3G). We did not observe a reduction in tumor-infiltrating γδ/αβ cell ratio or in the frequency of Vγ7$^+$ cells in tamoxifen-treated i*Trdc*$^{Scl2a1fl/+}$ mice (fig. S3H), which overall reinforces the notion that Glut1 does not play a major role in γδ T cell maintenance but is required for epithelium-resident γδ T cell function (*3*). Inactivation of Glut1 in γδ T cells did not affect the frequency of IFN-γ production by TCRαβ$^+$CD8αβ$^+$ cells (fig. S3I). By contrast, late Glut1 targeting (after the second DSS cycle) had no effect on tumor load or tumor-infiltrating γδ T cells (fig. S3, J to N). These results suggest that epithelium-resident γδ T cells exert immunosurveillance against

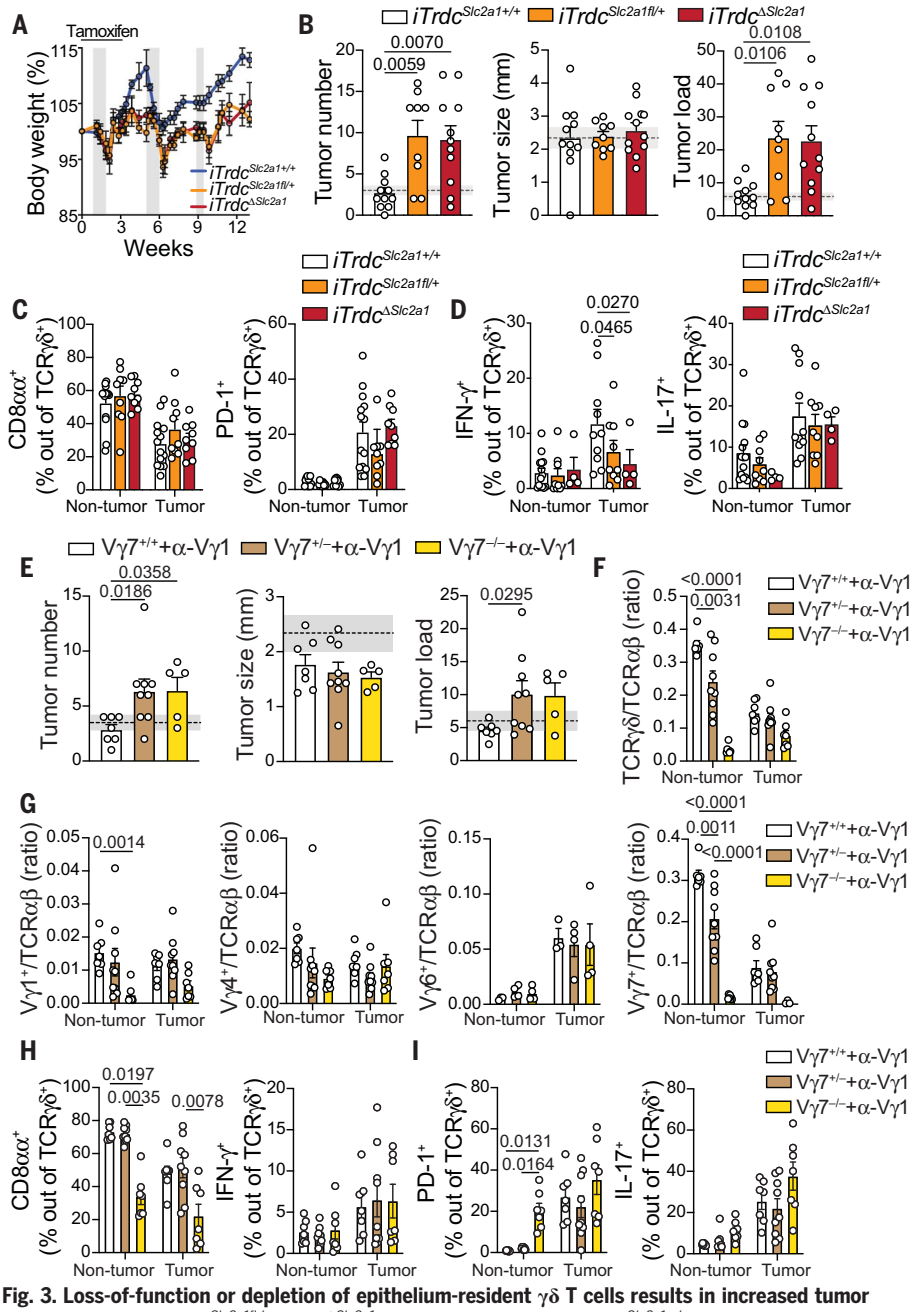


Fig. 3. Loss-of-function or depletion of epithelium-resident γδ T cells results in increased tumor numbers. (A to D) i*Trdc*$^{Slc2a1fl/+}$, i*Trdc*$^{\Delta Slc2a1}$, and littermate control (i*Trdc*$^{Slc2a1+/+}$) mice were subjected to AOM-DSS treatment, and tamoxifen was administered twice a week starting 1 week before until 2 weeks after AOM injection. Animals were analyzed 12 weeks after initial AOM injection. (**A**) Mean percentage of body weight changes during AOM-DSS treatment. Gray bars represent DSS treatment. (**B**) Tumor number, size, and load. Shaded areas bounded by dashed lines indicate means ± SEMs of all control C57BL6/J mice analyzed in fig. S3B (AOM-DSS model). (**C** and **D**) Flow cytometry analysis of γδ T cells from tumor or nontumor colonic tissue. Frequency of CD8α$^+$ (left) and PD-1$^+$ (right) (C) and IFN-γ$^+$ (left) and IL-17$^+$ (right) (D) among TCRγδ$^+$ cells in tumor or nontumor colonic tissue. (E to I) Vγ7$^{-/-}$, Vγ7$^{+/-}$, and littermate control mice (Vγ7$^{+/+}$) were subjected to the AOM-DSS model and analyzed 12 weeks after initial AOM injection. All groups were treated with 200 μg of anti-Vγ1–depleting antibody (2.11) twice a week, starting 1 week before AOM administration until the second DSS cycle. (**E**) Tumor number, size, and load. (**F** and **G**) Ratio of TCRγδ/αβ (F) and Vγ1/TCRαβ, Vγ4/TCRαβ, Vγ6/TCRαβ, and Vγ7/TCRαβ (G) among CD45$^+$ cells from colonic tissue. (**H** and **I**) Frequency of CD8α$^+$ (left) and IFN-γ$^+$ (right) (H) and PD-1$^+$ (left) and IL-17$^+$ (right) (I) among TCRγδ$^+$ cells. i*Trdc*$^{\Delta Slc2a1}$ data are pooled from three experiments with three to five animals per group. Vγ7$^{-/-}$ data are pooled from two experiments with three to five animals per group. For cytokine staining, cells were stimulated with PMA and ionomycin. Statistical *P* value differences are indicated. Statistics are by one-way ANOVA with Dunnett's multiple comparison test [(B) to (I)]. Error bars indicate SEMs.

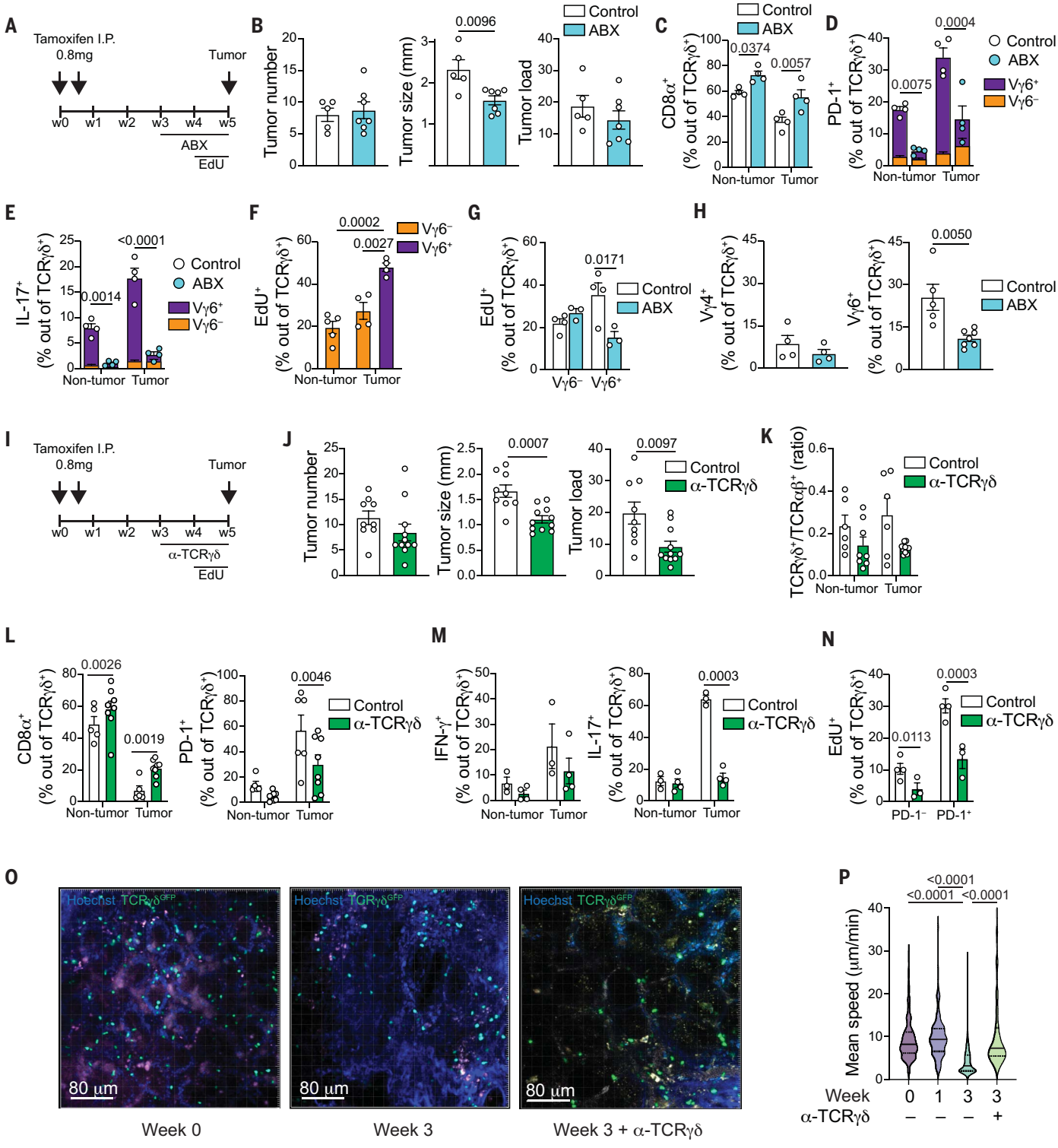


Fig. 4. Tumor-infiltrating IL-17$^+$ γδ T cells induce tumor growth in a microbiota- and TCR-dependent manner. *iCdx2*$^{\Delta APC}$ mice were treated with two intraperitoneal (I.P.) injections of 0.8 mg of tamoxifen and analyzed 5 weeks after [(A) to (N)] or at the indicated time [(O) and (P)]. For recovery and visualization of TCRγδ$^+$ cells, *iCdx2*$^{\Delta APC}$*Trdc*^{GFP} reporter mice were used [(I) to (P)]. Mice were treated with antibiotic mix (ABX) in the drinking water [(A) to (H)] or treated twice a week with 400 μg of anti-TCRγδ blocking antibody (UC7-13D5) for the last 2 weeks of the experiment [(I) to (N)]. For in vivo quantification of cell proliferation, animals were treated with EdU in the drinking water for 1 week before analysis. (**A** and **I**) Protocol. (**B** and **J**) Tumor number, size, and load. (C to H and K to N) Flow cytometry analysis of γδ T cells from tumor or nontumor colonic tissue. (**C**) Frequency of CD8α$^+$ cells among TCRγδ$^+$ cells. (**D** and **E**) Frequency of PD-1$^+$ (D) and IL-17$^+$ (E) among TCRγδ$^+$ cells. Vγ6$^+$ (purple) versus Vγ6$^-$ (orange) contribution to PD-1$^+$– and IL-17–producing γδ T cells is also shown. (**F** and **G**) Frequency of EdU incorporation by Vγ6$^-$ or Vγ6$^+$ among TCRγδ$^+$ cells. (G) shows tumor-infiltrating cells. (**H**) Frequency of Vγ4$^+$ (left) and Vγ6$^+$ (right) among TCRγδ$^+$ cells. (**K**) γδ^{GFP+}/TCRαβ ratio among CD45$^+$ cells. (**L** and **M**) Frequency of CD8α$^+$ (left) and

PD-1$^+$ (right) (L) and IFN-γ$^+$ (left) and IL-17$^+$ (right) (M) among γδ^{GFP+} cells. (**N**) Frequency of EdU incorporation by PD-1$^-$ or PD-1$^+$ among γδ^{GFP+} cells. (**O** and **P**) Intravital imaging of colonic γδ^{GFP+} cells. Animals were treated with α-TCRγδ blocking antibody (UC7-13D5) for 1 week before intravital imaging. (O) Representative images of γδ^{GFP+} cells before and 3 weeks after tamoxifen administration. Cells were tracked using *Imaris* (Bitplane AG) software. (P) Mean speed of individual tracks. Data from antibiotic-treated (ABX) i*Cdx2*$^{\Delta APC}$ mice are representative of three independent experiments with three to four animals per group. Data from i*Cdx2*$^{\Delta APC}$*Trdc*^{GFP} mice treated with UC7-13D5 are pooled from three independent experiments with three to four animals per group. Data from intravital imaging are representative of two experiments with two animals per group. Statistics are by one-way ANOVA with Dunnett's multiple comparison test [(C) to (G) and (K) to (N)], two-tailed *t* test [(B), (H), and (J)], and Kruskal-Wallis test with Benjamin multiple comparison test (P). For cytokine staining, cells were stimulated with PMA and ionomycin. Statistical *P* value differences are indicated. Error bars indicate SEMs.

epithelial tumors, and early functional impairment of Glut1-dependent γδ T cells favors CRC development.

To directly access an antitumor role of epithelium-resident γδ T cells, we generated Vγ7$^{-/-}$ mice by CRISPR-Cas9 genome editing (fig. S4, A to E). At steady state, Vγ7$^{-/-}$ mice display no changes in the tissues analyzed besides the intestine, which shows an overall reduction of γδ T cells; a decrease was also observed, albeit to a lesser degree, in Vγ7$^{+/-}$ mice, which suggests allelic exclusion by the unproductively rearranged TCR (fig. S4, B to E). TCRαβ$^+$ T cells showed no differences in the organs analyzed, except for an increase in frequency of TCRαβ$^+$CD8αα$^+$ natural IELs, suggesting a compensatory mechanism in the reduction of γδ IELs (fig. S4C). Vγ7$^{-/-}$ mice subjected to AOM-DSS treatment did not show differences in tumor number, size, and burden when compared with Vγ7$^{+/-}$ or Vγ7$^{+/+}$ littermate controls (fig. S5, A to G). Moreover, although there was a significant reduction in TCRγδ cells in nontumor areas of Vγ7$^{-/-}$ mice (assessed by the ratio of TCRγδ$^+$ to TCRαβ$^+$), the frequencies of IFN-γ, PD-1, and IL-17 expression by γδ T cells were similar between Vγ7$^{-/-}$ and Vγ7$^{+/-}$ littermate controls (fig. S5, C to E). Finally, tumor-infiltrating TCRαβ$^+$CD4$^+$ and TCRαβ$^+$CD8α$^+$ cells did not show differences in cytokine production between Vγ7$^{-/-}$ mice and Vγ7$^{+/-}$ littermate controls (fig. S5G). To address a possible compensatory role of Vγ1$^+$ γδ T cells in antitumor immunity in the absence of Vγ7$^+$ cells, we administered α-Vγ1–depleting antibodies (clone 2.11) starting 1 week before AOM-DSS treatment until the second DSS treatment to Vγ7$^{-/-}$ mice, Vγ7$^{+/-}$ mice, and littermate controls (Fig. 3, E to I). Both Vγ7$^{-/-}$ and Vγ7$^{+/-}$ mice treated with α-Vγ1–depleting antibodies showed increased tumor number and burden (Vγ7$^{+/-}$ group only) compared with treated littermate controls (Fig. 3E). Again, we observed a significant reduction in the ratio of TCRγδ$^+$ to TCRαβ$^+$ cells in the tumor for both Vγ7$^{+/-}$ and Vγ7$^{-/-}$ mice compared with littermate controls (Fig. 3F). Additionally, we did not observe differences in tumor-infiltrating Vγ subsets besides Vγ7$^+$ cells, which suggests that antibody-mediated depletion of Vγ1 cells was restricted to early stages of tumor progression (Fig. 3G). Nevertheless, we observed a reduced frequency of CD8αα-expressing γδ T cells in both tumor and nontumor areas and an increased frequency of PD-1$^+$ cells in nontumor areas in Vγ1-depleted Vγ7$^{-/-}$ mice compared with littermate controls (Fig. 3, H and I). We did not observe differences in IFN-γ or IL-17 production by tumor-infiltrating γδ T cells (Fig. 3, H and I). Tumor-infiltrating TCRαβ$^+$CD4$^+$ and TCRαβ$^+$CD8α$^+$ cells did not show differences in cytokine production between the groups (fig. S5H). Together, these data suggest that murine epithelium-resident Vγ1$^+$ and Vγ7$^+$ intestinal γδ T cells are important in the early control of tumor formation.

Protumor role of infiltrating murine Vγ4$^+$ and Vγ6$^+$ γδ T cells

Although our observations point to an important antitumor function for epithelium-resident γδ T cells, previous studies have also described that the tumor microenvironment, or the microbiome, can influence γδ T cell subsets to have an opposite, protumorigenic role during cancer progression (*3*). We therefore questioned whether the clonally expanded Vγ6$^+$ PD-1$^+$ IL-17–producing γδ T cells that accumulate among tumor-infiltrating lymphocytes contribute to CRC progression. We conditionally deleted Rorγt, which is the main transcription factor linked to IL-17 production in T cells (*32*), using the same strategy described above. To specifically target γδ T cells that accumulate during CRC progression, we treated i*Trdc*$^{\Delta Rorc}$ and littermate control mice with tamoxifen after the second DSS cycle (fig. S6, A to H). Although no differences in tumor numbers or burden were noted, late Rorγt deletion in γδ T cells results in smaller tumors (fig. S6B). Accordingly, this strategy leads to altered Vγ usage among tumor-infiltrating γδ T cells: Tamoxifen-treated i*Trdc*$^{\Delta Rorc}$ mice display significantly decreased Vγ6$^+$ and Vγ4$^+$ populations when compared with littermate controls (fig. S6C). Tamoxifen-treated i*Trdc*$^{\Delta Rorc}$ mice do not display changes in CD8αα-expressing γδ T cells (fig. S6D). However, late Rorγt targeting results in reduced frequency of PD-1$^+$– and IL-17–producing (71.2 and 39.08% suppression, respectively) tumor-infiltrating γδ T cells, affecting both Vγ6$^+$ and Vγ4$^+$ populations (fig. S6, C to E). Overall, we observed a 55% reduction in tumor-infiltrating Rorγt$^+$ γδ T cells in i*Trdc*$^{\Delta Rorc}$ mice after tamoxifen treatment when compared with littermate controls (fig. S6E). No difference was noted in Rorγt expression or IL-17 production among CD4$^+$ T cells (fig. S6F). Tumor-infiltrating CD8αβ$^+$ T cells displayed increased IFN-γ production in i*Trdc*$^{\Delta Rorc}$ mice when compared with littermate controls (fig. S6G). IL-17 and additional proinflammatory cytokines have been shown to promote tumor growth through the recruitment of monocytic myeloid-derived suppressor cells (M-MDSCs), inflammatory monocytes (both gated as CD11b$^+$Gr-1^{int}) and neutrophils, or granulocytic MDSCs (G-MDSCs) (both gated as CD11b$^+$Gr-1^{high}) (*14*, *16*, *17*, *19*, *33*–*35*). Accordingly, we observed a significant decrease in tumor-infiltrating CD11b$^+$Gr-1^{int} cells in i*Trdc*$^{\Delta Rorc}$ mice when compared with littermate controls (fig. S6H). These data indicate that Rorγt expression by γδ T cells is an important factor for the accumulation of IL-17–producing Vγ4$^+$ and Vγ6$^+$ TCRγδ$^+$ cells, which in turn contribute to a protumorigenic microenvironment.

Changes in resident bacterial communities have been associated with tumor burden in CRC models (*14*) and were shown to affect protumorigenic T cells, including Rorγt$^+$ IL-17–producing Vγ6$^+$ γδ T cells, in both lung and ovarian cancer models (*33*, *36*). Because of complex changes in gut microbiota composition after cycles of DSS treatment (*36*), we focused on the impact of microbiota changes and antibiotic treatment in the APC loss model. 16*S* ribosomal RNA (rRNA) sequencing from feces of mice subjected to the APC loss model revealed a sharp decrease in microbial diversity as well as broad changes in bacterial composition that can be detected as early as 2 weeks after tamoxifen administration and continue in the following time points (fig. S7, A and B). Consistent with a role of gut microbiota in promoting tumor growth (*14*), subjecting tamoxifen-treated i*Cdx2*$^{\Delta APC}$ mice to a broad-spectrum antibiotic cocktail (ABX; ampicillin, vancomycin, metronidazole, and neomycin) for only 1 week results in smaller tumors while not affecting tumor number or burden, which resembles changes observed upon late Rorγt targeting in the CAC model (Fig. 4, A and B). CD8αα$^+$ γδ T cells, primarily represented by Vγ1$^+$ and Vγ7$^+$ cells, display increased frequency in both nontumor and tumor sites in mice receiving ABX treatment (Fig. 4C). Moreover, ABX treatment results in reduced

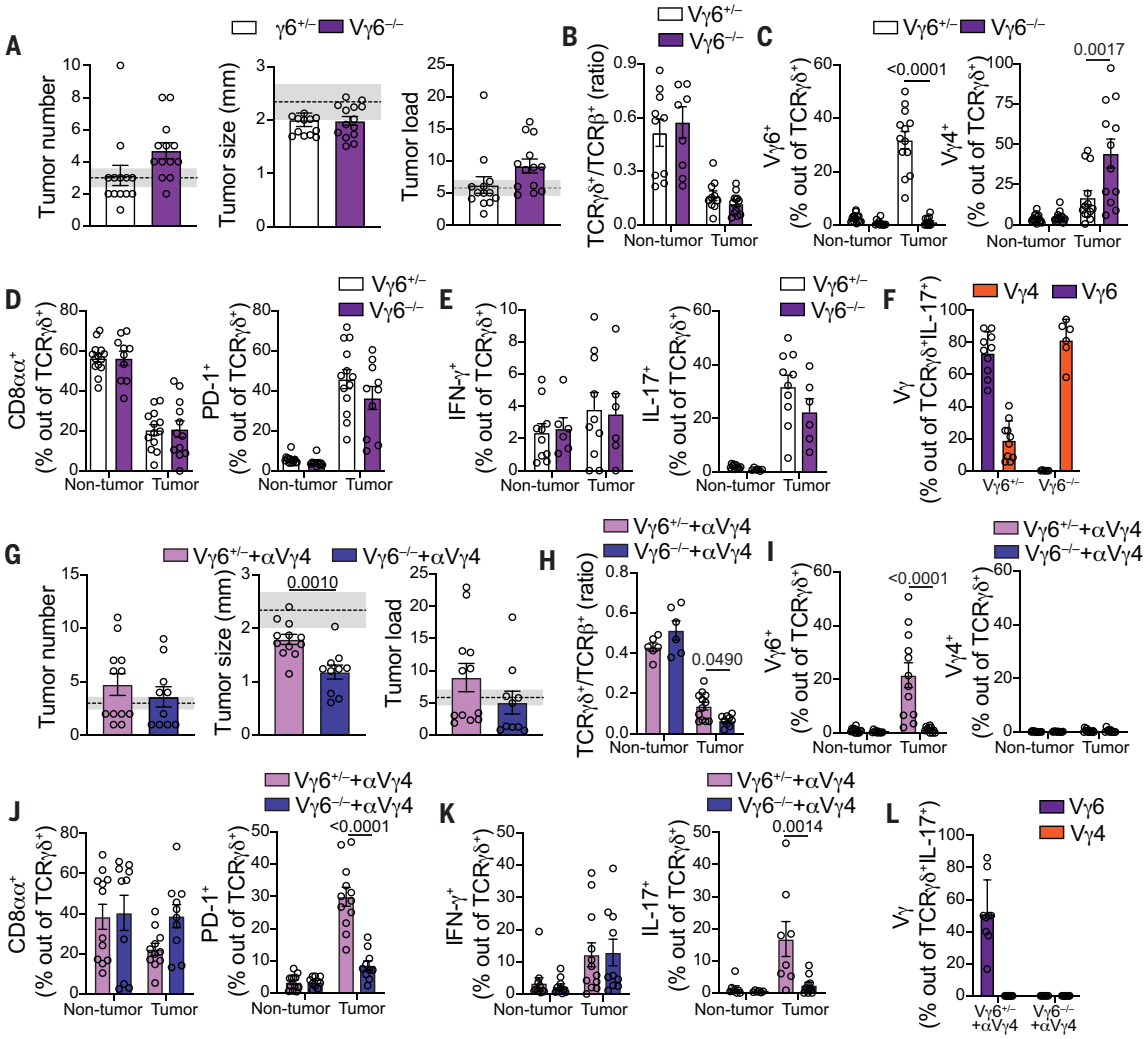


Fig. 5. Redundant tumor-infiltrating IL-17–producing Vγ6$^+$ and Vγ4$^+$ γδ cells promote tumor growth. (A to L) Female Vγ6$^{-/-}$ and Vγ6$^{+/-}$ littermate control mice were subjected to the AOM-DSS protocol and analyzed 12 weeks after initial AOM injection. In (G) to (L), mice received injections of 200 μg of α-Vγ4–depleting antibody (UC3-10A6) twice a week starting 1 week after the second DSS cycle (last 6 weeks of experiment). (**A** and **G**) Tumor number, size, and load. Shaded areas bounded by dashed lines indicate means ± SEMs of all control C57BL6/J mice analyzed in fig. S3B (AOM + DSS model). (B to F and H to L) Flow cytometry analysis of γδ T cells from tumor or nontumor colonic tissue. (**B** and **H**) TCRγδ/αβ ratio among CD45$^+$ cells from colonic tumor tissue. (**C** and **I**) Frequency of Vγ6$^+$ (left) and Vγ4$^+$ (right) among TCRγδ$^+$ cells. (**D** and **J**) Frequency of CD8α$^+$ (left) and PD-1$^+$ (right) among TCRγδ$^+$ cells. (**E** and **K**) Frequency of IFN-γ$^+$ (left) and IL-17$^+$ (right) among TCRγδ$^+$ cells. (**F** and **L**) Frequency of tumor-infiltrating Vγ6$^+$ and Vγ4$^+$ among IL-17–producing TCRγδ$^+$ T cells. Data from Vγ6$^{-/-}$ and α-Vγ4–treated Vγ6$^{-/-}$ are pooled from two and three experiments, respectively, with three to six animals per group. For cytokine staining, cells were stimulated with PMA and ionomycin. Statistical *P* value differences are indicated. Statistics are by one-way ANOVA with Dunnett's multiple comparison test [(C) to (F) and (I) to (L)] and two-tailed *t* test [(A), (B), (G), and (H)]. Error bars indicate SEMs.

frequency of IL-17– and PD-1–expressing γδ T cells (both >75% suppression), primarily within the Vγ6$^+$ subset (Fig. 4, D and E), even though Vγ6$^-$ γδ T cells (mostly composed of Vγ4$^+$) also contribute to PD-1 and IL-17 expression in tumor sites (fig. S2Q). To address whether Vγ6$^+$ cells accumulate during CRC progression as a result of increased recruitment or in situ proliferation, we performed in vivo 5-ethynyl-2′-deoxyuridine (EdU) labeling. We observe higher proliferation rates in Vγ6$^+$ versus Vγ6$^-$ γδ T cells, which are reduced (56.7%) upon ABX treatment (Fig. 4, F and G). Although microbiota depletion in tamoxifen-treated i*Cdx2*$^{\Delta APC}$ mice led to a decrease in intratumoral Vγ6$^+$ γδ T cells, it did not affect Vγ4$^+$ cells (Fig. 4H). Additionally, contrary to what was observed upon late Rorγt targeting in the CAC model, we did not observe changes in frequency of IFN-γ–producing CD8αβ$^+$ T cells or CD11b$^+$Gr-1$^+$ (both high and intermediate) cells in tumor areas of ABX-treated animals (fig. S7, C and D). These observations point to distinct susceptibility after microbiota manipulations by two main subsets of tumor-accumulating γδ T cells. Specifically, Vγ6$^+$, but not Vγ4$^+$, cells proliferate in response to microbiota and depend on microbiota signals to sustain PD-1 and IL-17 expression as well as to boost tumor growth.

Because their temporal dynamics and distinct functional properties were segregated on the basis of TCR V usage, we investigated a possible role for TCR engagement by epithelium-resident versus tumor-infiltrating γδ T cells. We treated i*Cdx2*$^{\Delta APC}$ mice with anti-TCRγδ blocking (nondepleting) antibodies (UC7-13D5) for 2 weeks starting 7 days before tamoxifen administration, therefore targeting TCRγδ-mediated signaling primarily on epithelium-resident Vγ1$^+$ and Vγ7$^+$ cells (fig. S7E). Early TCRγδ blockade did not affect tumor number

or size (fig. S7F); γδ/αβ ratio (fig. S7G); or frequencies of CD8α$^+$, PD-1$^+$, and IL-17– and IFN-γ–producing γδ T cells compared with those in control mice (fig. S7, H and J). Additionally, although tumor-infiltrating PD-1$^+$ cells show higher proliferation rates compared with PD-1$^-$ counterparts, early anti-TCRγδ treatment did not affect proliferation rates (fig. S7I). TCRγδ blocking did not affect CD4$^+$ T cell proliferation (fig. S7K).

Next, we treated i*Cdx2*$^{\Delta APC}$ mice with anti-TCRγδ blocking antibody for 2 weeks starting 3 weeks after initial tamoxifen treatment, hence also targeting TCRγδ-mediated signaling on tumor-infiltrating Vγ4$^+$ and Vγ6$^+$ cells (Fig. 4I). Late TCRγδ blockade led to a significant decrease in tumor size and burden but did not change tumor number (Fig. 4J). The γδ/αβ ratio, assessed by *Trdc*^{GFP}-driven green fluorescent protein (GFP) signals, was reduced in tumor areas, whereas the frequency of CD8αα$^+$ γδ T cells both in tumor and adjacent areas was increased (Fig. 4, K and L). Consistent with a preferential effect of late TCRγδ blockage on tumor-infiltrating Vγ4$^+$ and Vγ6$^+$ cells, we found a significant decrease in both PD-1$^+$– and IL-17–producing γδ T cells in tumor areas (Fig. 4, L and M). TCR blockade did not lead to changes in the frequency of IFN-γ–producing CD8αβ$^+$ T cells or CD11b$^+$Gr-1$^+$ (both high and intermediate) cells (fig. S7, L and M). In vivo EdU labeling confirmed that the heightened PD-1$^+$ γδ T cell proliferation rate in tumor areas is TCR dependent (Fig. 4N). Late TCRγδ blocking did not affect the proliferation of CD4$^+$ T cells (fig. S7N).

TCR engagement is associated with T cell movement arrest (*37*). To investigate whether the apparent TCR-dependent proliferation and protumorigenic function of tumor-infiltrating γδ T cell subsets correlate with cell motility changes, we performed live intravital multiphoton microscopy on cells from i*Cdx2*$^{\Delta APC}$*Trdc*^{GFP} reporter mice immediately before and 1 and 3 weeks after tamoxifen treatment (Fig. 4, O and P). Although we were unable to ascertain tumor borders or pretumor regions in the time points analyzed, total T cell displacement was overall consistent along time points and between cecal and colonic areas. Compared with videos obtained in mice either before or 1 week after tamoxifen treatment, colonic γδ T cells displayed reduced speed at 3 weeks after tamoxifen, suggesting ongoing TCR engagement (Fig. 4P and movies S1 and S2). Notably, this time point coincides with accumulation of Vγ6$^+$ cells in the colon of i*Cdx2*$^{\Delta APC}$ mice after tamoxifen treatment (Fig. 2G). Consistent with this notion, TCR blockade with in vivo anti-TCRγδ antibody treatment rescued cell velocity to the levels observed before or early after APC loss (Fig. 4P and fig. S7O). These results suggest that epithelium-resident antitumor γδ T cell subsets may function in a TCR-independent manner, whereas the γδ T cell subsets that accumulate during tumor progression function to boost tumor growth through the TCR.

Our findings so far indicate that IL-17–producing or PD-1–expressing tumor-infiltrating γδ T cells are composed ~70 to 85% of microbiota-dependent Vγ6$^+$ and ~10 to 20% of Vγ4$^+$ cells. To directly assess the roles of Vγ4$^+$ and Vγ6$^+$ γδ T cells in CRC progression, we generated Vγ4$^{-/-}$ and Vγ6$^{-/-}$ mice by CRISPR targeting of *Trgv4* and *Trgv6* genes, respectively (fig. S8, A to E, and fig. S9, A to E). Fully backcrossed Vγ4$^{-/-}$ mice were subjected to AOM-DSS and show no differences in tumor number, tumor size, or tumor-infiltrating γδ T cells (aside from Vγ4) when compared with heterozygous littermate controls (fig. S10, A to G). The ratio of γδ/αβ T cells among of tumor-infiltrating lymphocytes in Vγ4$^{-/-}$ mice remains the same as the ratio observed in Vγ4$^{+/-}$ controls (fig. S10B), which suggests a compensatory expansion of remaining γδ T cells, including Vγ6$^+$ T cells, in the absence of the Vγ4$^+$ subset.

We next analyzed B6-backcrossed naïve Vγ6$^{-/-}$ mice, which display a decreased γδ/αβ T cell ratio in the fat tissue at steady state, suggesting a lack of compensatory expansion of Vγ4$^+$ cells in noninflammatory settings (fig. S9B). Like Vγ4$^{-/-}$ mice, Vγ6$^{-/-}$ mice subjected to AOM-DSS displayed similar CRC development, progression (Fig. 5A), and parameters of tumor-infiltrating γδ T cells to those seen in Vγ6$^{+/-}$ littermate controls (Fig. 5, B to F). The absence of an otherwise large intratumor Vγ6$^+$ population did not result in an altered γδ/αβ T cell ratio (Fig. 5B), explained by a compensatory increase in other γδ T cells, particularly Vγ4$^+$ cells in tumor settings (Fig. 5C). In the absence of Vγ6$^+$ cells, we observed a sharp increase in tumor-infiltrating Vγ4$^+$ cells producing IL-17 (Fig. 5F). To address whether deletion of Vγ6 led to biased clonal expansion in the remaining cells, we performed scTCR-seq in Vγ4$^+$ cells in Vγ6$^{-/-}$ mice subjected to AOM-DSS (data S3). We identified large clonal expansions in Vγ6$^{-/-}$ mice, albeit similar to those observed in littermate control mice, both in tumor and nontumor areas (fig. S10, H and I). We did not observe differences in tumor-infiltrating CD4$^+$ T cells between Vγ6$^{-/-}$ and Vγ6$^{+/-}$ mice (fig. S10J). The overall similar tumor development and composition of tumor-infiltrating γδ T cells in Vγ4$^{-/-}$ and Vγ6$^{-/-}$ mice raised the possibility that tumor-infiltrating, PD-1$^+$ IL-17–producing Vγ4$^+$ and Vγ6$^+$ populations play redundant roles in promoting CRC growth.

To address possible compensatory and redundant roles between tumor-infiltrating Vγ4$^+$ and Vγ6$^+$ γδ T cells, we treated Vγ6$^{-/-}$ and Vγ6$^{+/-}$ littermate control mice with depleting anti-Vγ4 antibody (UC3-10A6) starting after the second DSS cycle until analysis. In contrast to untreated Vγ6$^{-/-}$ mice, Vγ4-depleted Vγ6$^{-/-}$ mice developed significantly smaller tumors than Vγ4-depleted Vγ6$^{+/-}$ littermate control mice, although no significant changes in tumor numbers or load were noted (Fig. 5G). In contrast to previous strategies, Vγ4-depleted Vγ6$^{-/-}$ mice display an ~50% reduction in tumor-infiltrating γδ/αβ T cell ratio (Fig. 5H). Vγ4-depleted Vγ6$^{-/-}$ mice also display enhanced intratumoral CD8αα$^+$ γδ T cells when compared with Vγ4-depleted Vγ6$^{+/-}$ mice (Fig. 5J), similar to what was observed in ABX- or α-TCRδγ–treated mice in the APC loss mouse model. Additionally, and consistent with a functional redundancy between Vγ4$^+$ and Vγ6$^+$ γδ T cells, Vγ4-depleted Vγ6$^{-/-}$ mice show ~80% reduction in the frequency of PD-1$^+$– and IL-17–producing γδ T cells within the tumor (Fig. 5, J to L). We did not observe changes in cytokine production by CD4$^+$ or CD8αβ$^+$ T cells (fig. S10, K and L) or in the frequency of CD11b$^+$Gr-1^{int} cells; however, Vγ4-depleted Vγ6$^{-/-}$ mice showed a significant decrease in CD11b$^+$Gr-1^{high} cells at tumor sites (fig. S10M), a phenotype likely also linked to the suppression of intratumoral IL-17 secretion by T cells. Hence, in sharp contrast to an antitumor role by epithelium-resident subsets, dominated by Vγ1$^+$ and Vγ7$^+$ γδ T cells, these data suggest redundant roles of tumor-infiltrating Vγ4$^+$ and Vγ6$^+$ γδ T cells in promoting CRC progression.

Discussion

Tumor-infiltrating lymphocytes are essential components of antitumor responses and represent major targets for immunotherapies (*6*, *11*, *19*, *38*). The cytokine and metabolic profiles of the γδ T cell subsets found in human and murine CRC implied opposing roles by γδ T cells found in the nontumor or steady-state epithelium versus γδ T cells that accumulate during cancer progression. Gain- and loss-of-function studies in mouse models confirmed that epithelial surveillance by steady-state IFN-γ–producing cytotoxic γδ populations helps prevent tumor initiation, whereas accumulating intratumor γδ T cells supports tumor progression.

Specialized dendritic epidermal T cell (DETC) γδ T cells in the skin, mostly composed of Vγ5$^+$ cells, have been shown to suppress tumor development through an NKG2D-dependent cytotoxic mechanism (*2*, *39*, *40*), whereas dermal γδ T cells, precommitted to IL-17 production and expressing Vγ4 or Vγ6, were shown to promote tumor growth (*19*). Although these observations are parallel to our findings, whether such anatomical segregation can also be observed in the intestinal epithelium versus the lamina propria in specific conditions (*27*) remains to be determined. Our observations in the intestine indicate that γδ IELs in pre- or nontumor areas harbor a diverse TCR repertoire while sharing primarily Vγ1 or Vγ7 segments,

particularly in the APC loss model. The accumulation of Vγ7$^+$ IELs was previously linked to binding of the germline Vγ7 chain to Btnl proteins—a tissue-specific selection not associated with the TCRγδ CDR3 region (*27*). Analogous to changes we observed in γδ T cells during CRC progression, previous studies described an irreversible expansion and repertoire reshaping of γδ T cells in chronically inflamed conditions, such as in celiac patients (*40*). Another important parallel with this study is their observation that γδ T cells from patients with active disease acquire a proinflammatory cytokine profile in contrast to γδ T cells isolated from patients in remission, which are primarily cytotoxic (*40*).

Our data collected from human specimens revealed expanded Vγ4Vδ1 and Vγ8Vδ3 epithelial-resident cells located in tumor-adjacent areas with strong cytotoxic gene signatures, suggesting antitumor activity paralleling murine Vγ1 and Vγ7. However, in contrast to a clear functional segregation of murine γδ T cells based on their Vγ usage, intratumor Vγ4Vδ1 cells displayed gene signatures associated with an intratumor metabolic adaptation and protumorigenic function. Along this line, a recent study on breast cancer proposed that tumor-infiltrating Vδ1$^+$ cells may have an immunosuppressive function (*41*). It is possible that this discrepancy between human and mouse data could be result of their disparate timing of CRC progression. A long-term exposure to the tumor microenvironment has been proposed to suppress an antitumor cytotoxic function by human gut–resident Vδ1$^+$ clones (*6*). Additionally, metabolic disorders have been linked to a decrease in tissue-resident γδ T cells during early tumorigenesis (*9*). Supporting the possibility of metabolic adaptation, a recent study has uncovered a metabolism-driven dichotomy in γδ T cell function: IL-17–producing γδ T cells were found to be dependent on oxidative phosphorylation, thriving in lipid-rich environments such as tumors, whereas IFN-γ–producing γδ T cells were found to require glycolysis for their energy expenditure (*31*). Conversely, previous observations have suggested a role for clonally restricted IL-17–producing T cells, including γδ cells, in tumor progression (*14*, *33*, *36*). These studies are in line with our observations that, like their response to invading bacteria (*3*), IFN-γ–producing γδ T cells, particularly the Vγ1$^+$ and Vγ7$^+$ subsets, display antitumor activity dependent on Glut1 expression, whereas IL-17–producing γδ T cells, particularly the Vγ6Vδ1 clone, were highly expanded in tumor areas.

Our study has several limitations. Although our results and previous literature (*14*, *16*, *17*, *19*, *33*, *34*) point to a tumor-progression role for intratumoral IL-17, which downstream mechanisms induced by IL-17—such as neutrophil and MDSCs recruitment—are necessary for the regulation of tumor growth remain to be defined. Because IL-17 production by γδ T cells in the gut has also been linked to tissue repair (*42*), it remains possible that in addition to IL-17, additional factors or molecules expressed by Vγ4$^+$ or Vγ6$^+$ cells aid tumor growth in the balance with tissue-repair mechanisms. It remains unclear whether high PD-1 expression, specific localization within the tumor, or other factors distinguish a tumor progression role for γδ T cells versus other IL-17–secreting cells, such as T_H17 cells. Additionally, our studies did not define the mechanisms by which the tumor microenvironment mediates the accumulation of microbiota-dependent Vγ6$^+$ and microbiota-independent Vγ4$^+$ subsets that aid tumor growth; it is possible that analogous mechanisms to the Btnl-dependent Vγ7 selection are used (*26*, *27*, *43*, *44*). In fact, a recent study has suggested that the expression of Btnl2 by tumor cells can specifically recruit protumorigenic IL-17–producing γδ T cells (*45*). Although the reduced number of patients and cells analyzed limit broader generalizations regarding public clones, TCR usage biases, and linked signatures, our analyses of infiltrating γδ T cells in human CRC provide an important parallel for the mechanistic details uncovered in murine studies. Studies using larger cohorts of CRC patients with variable tumor stages and analyzing a larger number of cells could further clarify pro- and antitumor activity by specific human γδ subsets. Nonetheless, our results caution against broad targeting of γδ T cells in future immune-therapy strategies, yet they also open possibilities of specific targeting of γδ T subsets based on V gene usage or their metabolic adaptation.

REFERENCES AND NOTES

1. J. Ettersperger *et al.*, *Immunity* **45**, 610–625 (2016).
2. T. Sujino *et al.*, *Science* **352**, 1581–1586 (2016).
3. D. P. Hoytema van Konijnenburg *et al.*, *Cell* **171**, 783–794.e13 (2017).
4. B. D. McDonald, B. Jabri, A. Bendelac, *Nat. Rev. Immunol.* **18**, 514–525 (2018).
5. K. L. Edelblum *et al.*, *Gastroenterology* **148**, 1417–1426 (2015).
6. J. Mikulak *et al.*, *JCI Insight* **4**, e125884 (2019).
7. W. Scheper, Z. Sebestyen, J. Kuball, *Front. Immunol.* **5**, 601 (2014).
8. B. Silva-Santos, K. Serre, H. Norell, *Nat. Rev. Immunol.* **15**, 683–691 (2015).
9. G. Tie *et al.*, *Cancer Res.* **77**, 2351–2362 (2017).
10. S. Matsuda, S. Kudoh, S. Katayama, *Jpn. J. Cancer Res.* **92**, 880–885 (2001).
11. Y. Wu *et al.*, *Sci. Transl. Med.* **11**, eaax9364 (2019).
12. S. Grivennikov *et al.*, *Cancer Cell* **15**, 103–113 (2009).
13. S. I. Grivennikov, F. R. Greten, M. Karin, *Cell* **140**, 883–899 (2010).
14. S. I. Grivennikov *et al.*, *Nature* **491**, 254–258 (2012).
15. O. Dmitrieva-Posocco *et al.*, *Immunity* **50**, 166–180.e7 (2019).
16. S. Ma *et al.*, *Cancer Res.* **74**, 1969–1982 (2014).
17. S. B. Coffelt *et al.*, *Nature* **522**, 345–348 (2015).
18. P. Wu *et al.*, *Immunity* **40**, 785–800 (2014).
19. B. Silva-Santos, S. Mensurado, S. B. Coffelt, *Nat. Rev. Cancer* **19**, 392–404 (2019).
20. J. P. Fisher *et al.*, *Clin. Cancer Res.* **20**, 5720–5732 (2014).
21. L. Tan *et al.*, *Cell Rep.* **27**, 3657–3671.e4 (2019).
22. I. Okayasu, T. Ohkusa, K. Kajiura, J. Kanno, S. Sakamoto, *Gut* **39**, 87–92 (1996).
23. Y. Feng *et al.*, *Am. J. Pathol.* **183**, 493–503 (2013).
24. T. Hinoi *et al.*, *Cancer Res.* **67**, 9721–9730 (2007).
25. R. Di Marco Barros *et al.*, *Cell* **167**, 203–218.e17 (2016).
26. A. Jandke *et al.*, *Nat. Commun.* **11**, 3769 (2020).
27. D. Melandri *et al.*, *Nat. Immunol.* **19**, 1352–1365 (2018).
28. F. R. Greten, S. I. Grivennikov, *Immunity* **51**, 27–41 (2019).
29. M. Yassin *et al.*, *Immunology* **158**, 35–46 (2019).
30. Y. Chen, K. Chou, E. Fuchs, W. L. Havran, R. Boismenu, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14338–14343 (2002).
31. N. Lopes *et al.*, *Nat. Immunol.* **22**, 179–192 (2021).
32. I. I. Ivanov *et al.*, *Cell* **126**, 1121–1133 (2006).
33. C. Jin *et al.*, *Cell* **176**, 998–1013.e16 (2019).
34. F. McAllister *et al.*, *Cancer Cell* **25**, 621–637 (2014).
35. M. F. Tavazoie *et al.*, *Cell* **172**, 825–840.e18 (2018).
36. M. Rei *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3562–E3570 (2014).
37. H. D. Moreau *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12151–12156 (2015).
38. M. Girardi *et al.*, *J. Exp. Med.* **198**, 747–755 (2003).
39. H. Cheroutre, F. Lambolez, D. Mucida, *Nat. Rev. Immunol.* **11**, 445–456 (2011).
40. T. Mayassi *et al.*, *Cell* **176**, 967–981.e19 (2019).
41. G. Chabab *et al.*, *J. Leukoc. Biol.* **107**, 1057–1067 (2020).
42. J. S. Lee *et al.*, *Immunity* **43**, 727–738 (2015).
43. A. C. Hayday, *J. Immunol.* **203**, 311–320 (2019).
44. S. P. Fahl *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1889–1894 (2018).
45. Y. Du *et al.*, *Nat. Commun.* **13**, 231 (2022).

ACKNOWLEDGMENTS

We thank all Mucida laboratory members and Rockefeller University employees for their continuous assistance and A. Rogoz and S. Gonzalez for the maintenance of mice. We thank R. O'Brien for the anti-Vγ7 hybridoma and Y. Belkaid for the anti-Vγ6 hybridoma. We also thank the Victora and Lafaille laboratories for fruitful discussions. **Funding:** This work was supported by National Institutes of Health (NIH) grant R01CA218133 (S.G.), NCTSA UL1TR001866 (Y.A.), the Howard Hughes Medical Institute (D.M.), NIH grant R01DK093674 (D.M.), NIH grant R01DK113375 (D.M.), the Food Allergy FARE/FASI Consortium (D.M.), the Mathers Foundation and Pershing Square Sohn Cancer Research Alliance (D.M.), and the Black Family Metastasis Foundation (B.S.R.). **Author contributions:** Conceptualization: B.S.R. and D.M. Data curation: T.B.R.C., B.S.R., P.W.D., and A.M.B. Formal analysis: B.S.R., P.W.D., A.L., and A.M.B. Funding acquisition: D.M., B.S.R., and S.G. Investigation: B.S.R., P.W.D., A.L., A.M.B., C.S.M., A.E.K., A.S.H., I.Z.K., V.S.S., O.E., C.Z., A.R., M.S., and J.B. Methodology: B.S.R., S.G., D.M., A.M.B., T.B.R.C., and J.B. Project administration: B.S.R. and D.M. Resources: A.S.H. and Y.A. Software: T.B.R.C. Supervision: D.M., B.S.R., and S.G. Validation: B.S.R. Visualization: B.S.R. and T.B.R.C. Writing – original draft: B.S.R. and D.M. Writing – review & editing: B.S.R. and D.M. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. RNA-seq data are available in the Gene Expression Omnibus under accession no. GSE205720. E8^{cre} mice were generated by D. Littman (NYU), were obtained under a material transfer agreement (no. RU 2851) with the La Jolla Institute for Allergy & Immunology in 2013, and are available from The Jackson Laboratory (008766). *Scl2a1*$^{fl/fl}$ were generated and provided by D. Abel under a material transfer agreement with the University of Iowa Research Foundation (no. 2016-0268 / RU 17-011) in 2016 and available from The Jackson Laboratory (031871).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj8695
Materials and Methods
Figs. S1 to S10
Tables S1 to S4
References (*46*–*60*)
MDAR Reproducibility Checklist
Movies S1 and S2
Data S1 to S3

View/request a protocol for this paper from *Bio-protocol*.

Submitted 7 June 2021; resubmitted 1 April 2022
Accepted 15 June 2022
10.1126/science.abj8695

ASTEROIDS

Spacecraft sample collection and subsurface excavation of asteroid (101955) Bennu

D. S. Lauretta[1]*, C. D. Adam[2], A. J. Allen[3], R.-L. Ballouz[1], O. S. Barnouin[4], K. J. Becker[1], T. Becker[1], C. A. Bennett[1], E. B. Bierhaus[5], B. J. Bos[6], R. D. Burns[6], H. Campins[3], Y. Cho[7], P. R. Christensen[8], E. C. A. Church[5], B. E. Clark[9], H. C. Connolly Jr.[10,1], M. G. Daly[11], D. N. DellaGiustina[1], C. Y. Drouet d'Aubigny[1], J. P. Emery[12], H. L. Enos[1], S. Freund Kasper[5], J. B. Garvin[6], K. Getzandanner[6], D. R. Golish[1], V. E. Hamilton[13], C. W. Hergenrother[1]†, H. H. Kaplan[6], L. P. Keller[14], E. J. Lessac-Chenen[2], A. J. Liounis[6], H. Ma[5], L. K. McCarthy[2], B. D. Miller[5], M. C. Moreau[2], T. Morota[7], D. S. Nelson[2], J. O. Nolau[3], R. Olds[5], M. Pajola[15], J. Y. Pelgrift[2], A. T. Polit[1], M. A. Ravine[16], D. C. Reuter[6], B. Rizk[1], B. Rozitis[17], A. J. Ryan[1], E. M. Sahr[2], N. Sakatani[18], J. A. Seabrook[11], S. H. Selznick[1]†, M. A. Skeen[5], A. A. Simon[6], S. Sugita[7], K. J. Walsh[13], M. M. Westermann[1], C. W. V. Wolner[1], K. Yumoto[7]

Carbonaceous asteroids, such as (101955) Bennu, preserve material from the early Solar System, including volatile compounds and organic molecules. We report spacecraft imaging and spectral data collected during and after retrieval of a sample from Bennu's surface. The sampling event mobilized rocks and dust into a debris plume, excavating a 9-meter-long elliptical crater. This exposed material is darker, spectrally redder, and more abundant in fine particulates than the original surface. The bulk density of the displaced subsurface material was 500 to 700 kilograms per cubic meter, which is about half that of the whole asteroid. Particulates that landed on instrument optics spectrally resemble aqueously altered carbonaceous meteorites. The spacecraft stored 250 ± 101 grams of material, which will be delivered to Earth in 2023.

Carbonaceous asteroids contain materials that are billions of years old and preserve a record of the earliest stages of Solar System evolution (*1*). Hydrated minerals and organic compounds in carbonaceous chondrite meteorites—which are thought to be fragments of these asteroids—indicate that they may have transported water and prebiotic organic molecules to Earth (*2*). However, meteorites must survive atmospheric entry, are exposed to terrestrial contamination, and lack geologic context. The Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer (OSIRIS-REx) mission has a primary goal to collect a pristine sample of surface material, from a well-characterized carbonaceous asteroid, and return it to Earth (*3*, *4*).

[1]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA. [2]KinetX, Simi Valley, CA, USA. [3]Department of Physics, University of Central Florida, Orlando, FL, USA. [4]The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA. [5]Lockheed Martin Space, Littleton, CO, USA. [6]Goddard Space Flight Center, Greenbelt, MD, USA. [7]Department of Earth and Planetary Environmental Science, University of Tokyo, Tokyo, Japan. [8]School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA. [9]Department of Physics and Astronomy, Ithaca College, Ithaca, NY, USA. [10]Department of Geology, Rowan University, Glassboro, NJ, USA. [11]Department of Earth and Space Science and Engineering, York University, Toronto, ON, Canada. [12]Department of Astronomy and Planetary Science, Northern Arizona University, Flagstaff, AZ, USA. [13]Southwest Research Institute, Boulder, CO, USA. [14]Johnson Space Center, Houston, TX, USA. [15]Astronomical Observatory of Padova, Italian National Institute for Astrophysics (INAF), Padova, Italy. [16]Malin Space Science Systems, San Diego, CA, USA. [17]School of Physical Sciences, Open University, Milton Keynes, UK. [18]Department of Physics, Rikkyo University, Tokyo, Japan.
*Corresponding author. Email: lauretta@arizona.edu
†Present address: Ascending Node Technologies, Tucson, AZ, USA.

OSIRIS-REx spent about 2 years surveying (101955) Bennu, a ~500-m-diameter carbonaceous rubble-pile asteroid [for example, (*4*–*6*)]. The microgravity environment (*7*) and unexpectedly rough terrain (*5*) posed challenges for collecting a sample (*4*). Nevertheless, after global observations of the surface, a site nicknamed Nightingale was chosen for sampling (*4*, *8*). We describe the sample collection process and its results.

Site of sample collection

Nightingale is situated within the 20-m-diameter Hokioi crater [center coordinates of about 56°, 43° (*9*)] (Fig. 1A). This site was selected on the basis of spacecraft navigation and safety considerations (*4*, *10*) and an expected higher abundance—relative to the generally boulder-dominated asteroid (*5*)—of surface particles <2 cm in diameter (*11*), the size class that is ingestible by the spacecraft's Touch-and-Go Sample Acquisition Mechanism (TAGSAM) (fig. S1) (*9*). Spectroscopic observations indicated that minerals present at Nightingale include hydrated phyllosilicates (*6*, *12*, *13*), the iron oxide magnetite (*5*, *12*, *14*), organic molecules (*6*, *15*, *16*), and carbonates (*6*, *17*). In panchromatic images, Nightingale exhibits a salt-and-pepper appearance (Fig. 1B), suggesting that Bennu's two primary lithologies—dark with low thermal inertia and brighter with higher thermal inertia (*14*, *18*)—might both be sampled. Hokioi crater is spectrally redder than the average surface of Bennu and thus is thought to be among the youngest features on the surface (*14*). The crater's midlatitude location limits the peak temperatures it experiences to ~360 K (versus ~390 K at Bennu's equator) (*19*), reducing the thermal processing of material there (*20*).

Sample collection operations and initial surface response

On 20 October 2020, the spacecraft left orbit and descended toward the surface of Bennu for sample collection, guided by autonomous optical navigation (fig. S2) (*9*). TAGSAM (fig. S1), which comprises a circular sample collection device connected to the spacecraft by a pogo stick–like arm, briefly contacted the asteroid before the spacecraft's back-away thrusters fired (fig. S2)—hence, the moniker Touch-and-Go (TAG) for the sampling maneuver. We investigated the sampling event using data acquired by the SamCam imager, which is part of the OSIRIS-REx Camera Suite (OCAMS); the NavCam 2 imager, which is part of the Touch-and-Go Camera System (TAGCAMS); and the spacecraft's inertial measurement unit (IMU) (*9*).

Contact and immediate disturbance

TAGSAM contacted the surface at latitude 55.8993°, longitude 41.8412°, with a downward velocity of 10.05 ± 0.004 cm s^{-1} (fig. S3), within 73 cm of the targeted location (*9*). Camera and IMU data indicate an initial surface contact force of between 10 and 15 N (*21*). One second after contact, TAGSAM released a jet of nitrogen gas to fluidize unconsolidated surface material and guide it into the collection chamber. About 6 s after contact, the spacecraft retained a downward velocity of ~4 cm s^{-1}. We attribute the velocity change of 6 cm s^{-1} to TAGSAM gas release. The spacecraft then fired eight of its 4.5-N thrusters, initiating the back-away maneuver. Three seconds of thruster firing arrested the residual downward velocity, then the spacecraft began to retreat from Bennu. By then, TAGSAM had penetrated 48.8 cm into the subsurface. The thrusters fired for a total of 25.7 s, with TAGSAM rising above the original surface height 16.6 s after initial contact. The spacecraft then drifted away from Bennu on a hyperbolic trajectory, at about 30 cm s^{-1}.

The surface of Bennu responded to contact as a compliant, viscous fluid, providing minimal resistance to the downward motion of the spacecraft (fig. S4 and movie S1) (*21*). This response is consistent with simulations of TAGSAM interacting with regolith (unconsolidated rocks and dust) that has near-zero interparticle cohesion (*21*–*23*) and low-gravity experiments in which loose material is easily mobilized (*24*).

Images collected by SamCam 0.8 ± 0.1 s after contact (0.32 s before gas firing) showed evidence of surface disturbance in all directions around TAGSAM (Fig. 1, C and D). We interpret fig. S5 and movie S2 as showing TAGSAM partially disrupting a ~20-cm-long

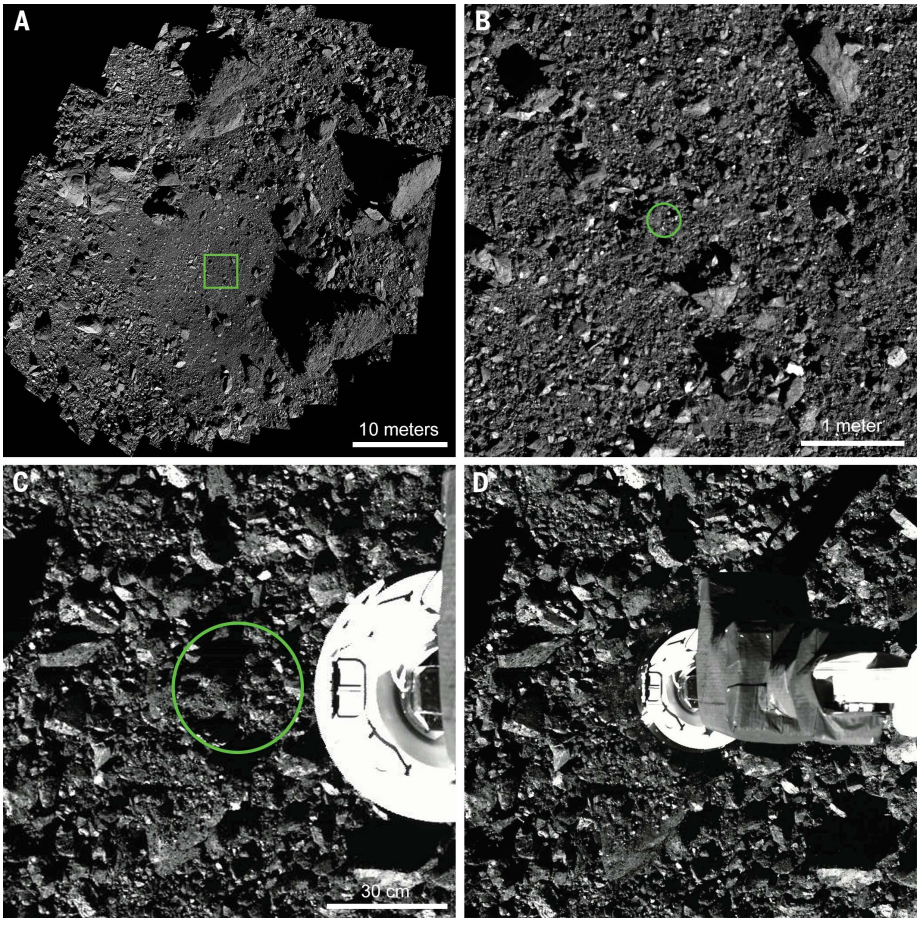


Fig. 1. The Nightingale sample site. (**A**) Hokioi crater, in a mosaic of images taken by the OCAMS PolyCam imager on 3 March 2020 with a pixel scale of 4 mm. The box indicates the area shown in (B). (**B**) PolyCam image taken 3 March 2020 with a pixel scale of 0.3 cm, showing a close-up of Hokioi crater. The green circle has a diameter of 32 cm, the same size as TAGSAM, indicating where surface contact occurred. (**C**) SamCam image of the TAGSAM contact point collected on 20 October 2020, 38 s before surface contact. The green circle is the same as in (B). The image has been cropped to show the same field of view as in (D). (**D**) SamCam image of TAGSAM contacting Bennu, collected 0.8 s after the first indication of surface interaction from the IMU, 0.32 s before gas bottle firing. Bennu north (+*z*) is up in all images.

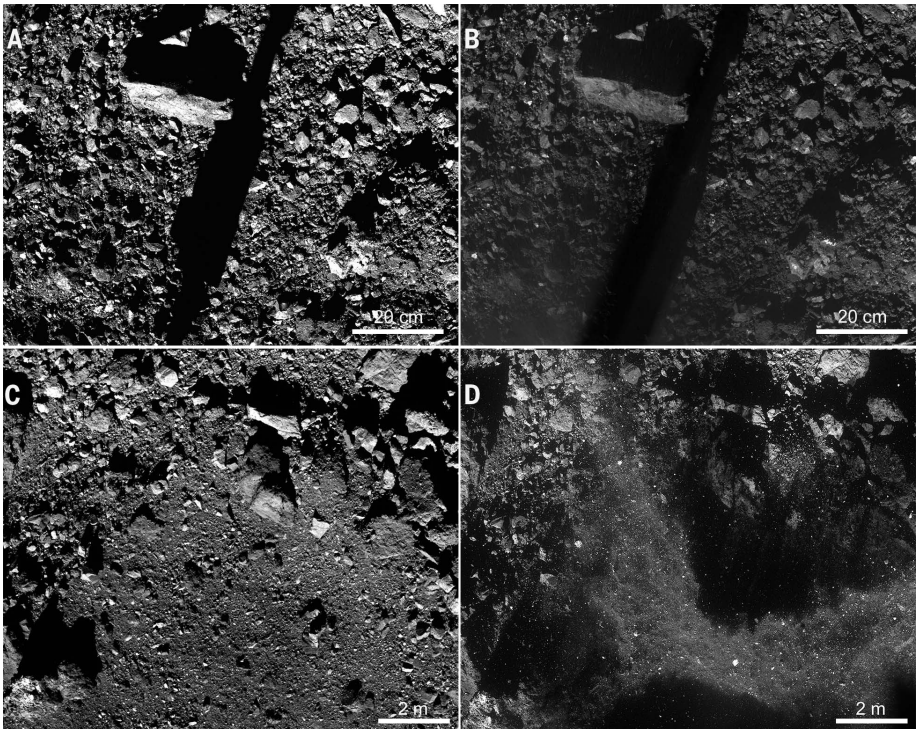


Fig. 2. Dust and debris mobilized by sampling. (**A**) NavCam 2 image taken immediately before surface contact, at 21:49:49 UTC on 20 October 2020. The shadow of the TAGSAM arm runs through the middle of the image. Pixel scale is 1 mm. (**B**) Post-contact NavCam 2 image taken at 21:49:50 UTC, with the same field of view as in (A), showing a cloud of submillimeter dust emanating from the TAGSAM contact point (below the field of view). Pixel scale is 1 mm. (**C**) Pre-contact, high-altitude NavCam2 image taken at 21:42:41 UTC. The TAGSAM contact point is out of the field of view, west of the scene. The image has been cropped to show a similar field of view as in (D). Pixel scale is 15 mm. (**D**) Post-contact, high-altitude NavCam 2 image taken at 21:51:13 UTC, showing the lofted debris plume. Pixel scale is 9 mm.

boulder near the eastern edge of its contact point. Another boulder, 40 cm long and to the southeast of TAGSAM, responded like a rigid plate, tilting up and launching small particles perched on its surface (fig. S6 and movie S3). These different responses support the inference (from remote sensing) that Bennu's boulders have a range of strengths (*18*, *25*).

Mobilization of dust

Comparing the NavCam 2 image acquired right before surface contact (Fig. 2A) to the image taken 0.1 s after gas release (Fig. 2B) shows a cloud of unresolved fine particulates emanating from the contact point. We used the NavCam 2 images in Fig. 2, A and B, to constrain the mass and particle sizes of this dust cloud. We measured the optical depth of the dust component and integrated over distance from the center of the outflow. The reduction in brightness between the two images constrains the optical model, which indicates average opacity of 0.47, generated by submillimeter particles (*9*).

We combined the opacity measurement with an estimated particle size frequency distribution (PSFD) with a maximum size of 3 mm (*9*), chosen on the basis of the cumulative PSFD from other populations of particles observed at Bennu (*11*, *26*–*28*) and a theoretical distribution (*9*). We derived a dust mass of 1.8 (−1.1/+1.2) kg. Assuming a particle density similar to that of carbonaceous chondrites (2000 kg m^{-3}), this mass is equivalent to a dust layer 3.2 (−2.2/+3.2) mm thick spread over a 60-cm-diameter circle (twice the diameter of TAGSAM). Previous studies had indicated little to no dust on Bennu's surface, with layer thicknesses tens of micrometers or less (*13*, *18*). Therefore, the subsurface must contain dust that was blown to the surface by the gas release.

Debris plume

Images collected during back-away show a high-opacity debris plume with a large spatial extent, smooth texture (relative to the unperturbed

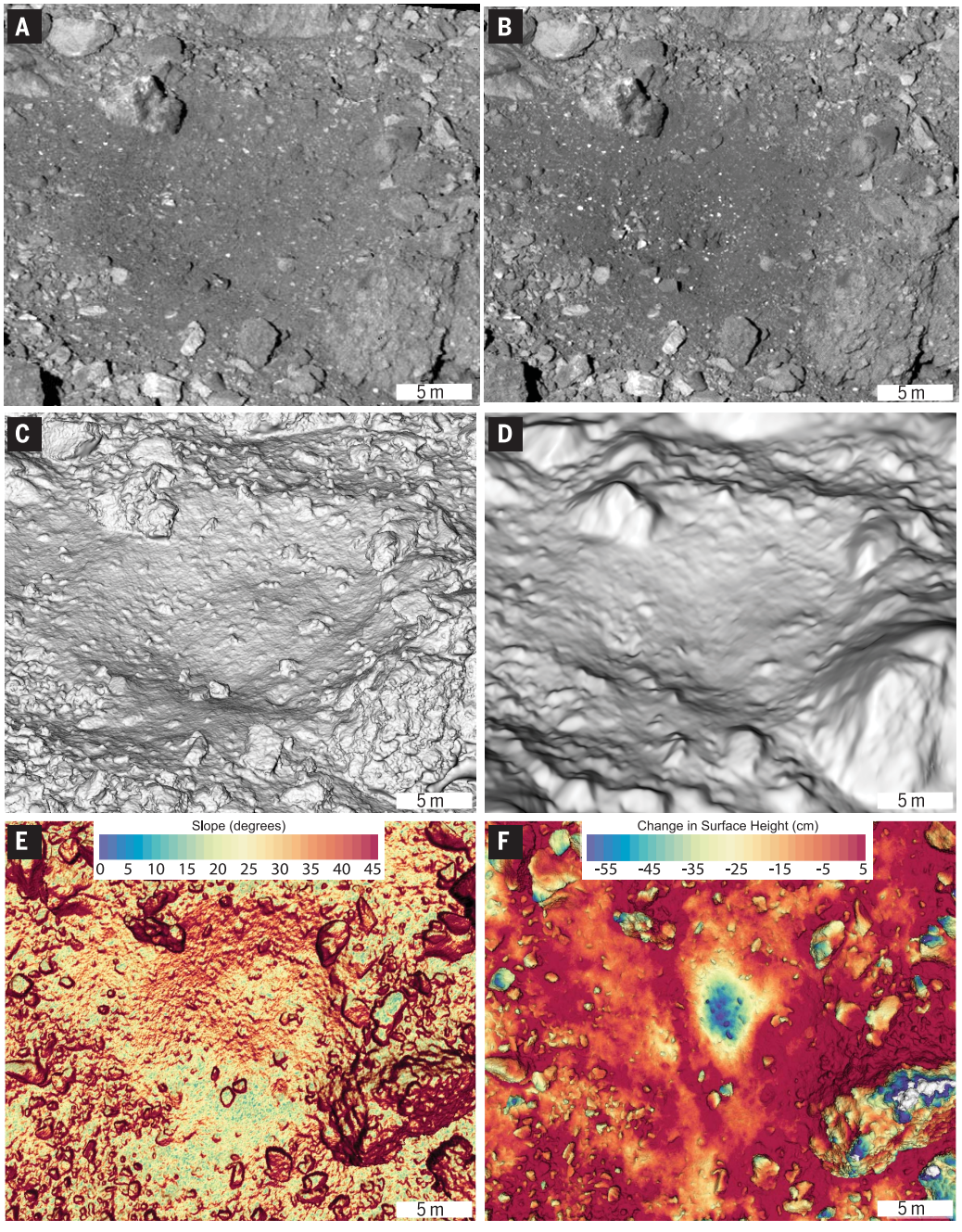


Fig. 3. Changes in surface appearance and topography after sampling. (**A**) Pre-sampling image collected on 7 March 2019. (**B**) Post-sampling image collected on 7 April 2021 (final flyby). These images have a pixel scale of about 5 cm. (**C**) Pre-sampling DTM based on OLA data, with a ground sample distance of 5 cm. (**D**) Post-sampling DTM constructed by using stereophotoclinometry (*9*), with a ground sample distance of 5 cm. The change in topography between (C) and (D) is real, whereas different surface textures in (C) and (D) are due to the different DTM construction methods. (**E**) Pre-sampling slope map, based on the OLA DTM and calculated surface accelerations (*7*). (**F**) Difference in height between the pre- and post-sampling DTMs (*9*). The deformation near the center is the TAG crater. All DTMs are viewed from an angle of ~68°, and Bennu north (+*z*) is up in all images.

surface), and well-defined shadows (Fig. 2, C and D, and fig. S7). Quantitative image analyses (*9*) indicate that this optically dense, lofted debris plume had a vertical and lateral velocity distinct from the asteroid surface. The morphology of the plume was due to a combination of subsurface excavation by TAGSAM gas release and pressure from the thruster exhaust. The thruster exhaust modified the debris plume, arresting the lateral velocity of finer particulates while permitting larger boulders to continue their eastward motion, indicated by shadows on a flat, bright boulder (movie S4). There was a steep gradient in dynamic pressure surrounding the thruster plumes, causing a range of effects, from total redirection of all particles to only affecting small particulates.

NavCam 2 images show an object in the plume colliding with a stationary rock on the surface, while a nearby tumbling rock spins almost in place, relative to Bennu (movies S4 and S5). The impact occurs at ~10 cm s^{-1}, breaking the impactor apart and changing its direction, before it disappears into the heavily shadowed region. PolyCam images collected during a later flyby (*9*) show that a 1.25-m boulder, which was directly under one set of thrusters, was transported 12 m to well outside Hokioi crater (fig. S8 and movie S6).

These observations support our interpretation that TAGSAM partially disrupted a boulder at the contact point. We suggest that some of the fine material in the debris plume was produced during the sampling event through the destruction of initially larger rocks. We measured the size of moving particles in the debris plume in consecutive NavCam 2 images (*9*). We found a power-law PSFD from 2.5 to 30 cm with an index of −3.3 ± 0.4, whereas previous high-resolution measurements of Nightingale found an index of −2.3 ± 0.3 (fig. S9 and table S1). This difference indicates that the material mobilized by sampling contained a higher proportion of fine particulates than the precontact surface.

Properties of the subsurface exposed by sampling

The spacecraft, which had been drifting away from the asteroid since sample collection, returned to Bennu on 7 April 2021 for a final flyby to characterize the newly exposed subsurface (fig. S10 and tables S2 and S3). Observations were taken 3.7 km from Bennu in a series of north-south linear scans, optimized for the PolyCam telescopic imager, which obtained images at 5 cm pixel scale (nadir resolution) (*9*). Observations were also taken by the MapCam multispectral imager at 25 cm pixel scale, the OSIRIS-REx Thermal Emission Spectrometer (OTES) at 28.2 m footprint, and the OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS) at 14.1 m footprint (*9*).

Crater formed by spacecraft interaction

Comparison of PolyCam images, collected during the global survey of Bennu in 2019 and the final flyby in 2021 (Fig. 3, A and B), shows changes to the Hokioi crater caused by the sampling event. We applied stereophotoclinometry to the final flyby images to construct a digital terrain model (DTM) with 5-cm ground sample distance (Fig. 3D) (*9*). We registered portions of the images to landmarks exterior to Hokioi crater that were not affected by sampling and compared the results with a previous DTM constructed at the same resolution, using OSIRIS-REx Laser Altimeter (OLA) data taken before sampling (Fig. 3C) (*9*). We found that the sampling event removed the edge of a preexisting debris apron at the base of the Hokioi crater wall (Fig. 3, A and C, and fig. S8) and replaced it with a crater (centered at 55.997°, 44.971°) with several boulders at its bottom, which we nickname the TAG crater (Fig. 3, B, D, and F; and figs. S8, S11, and S12). This excavated crater is elliptical (9.0 by 6.5 m), with its long axis oriented north to south. Its average lateral dimension is 7.8 ± 1.8 m, and its depth is 0.68 ± 0.1 m. The displaced volume is 12.2 ± 0.9 m^3. It is surrounded by decimeter-scale rocks that were transported by the sampling event into an arc (fig. S8), resembling the stone rings seen around small impact craters elsewhere on Bennu's surface (*29*).

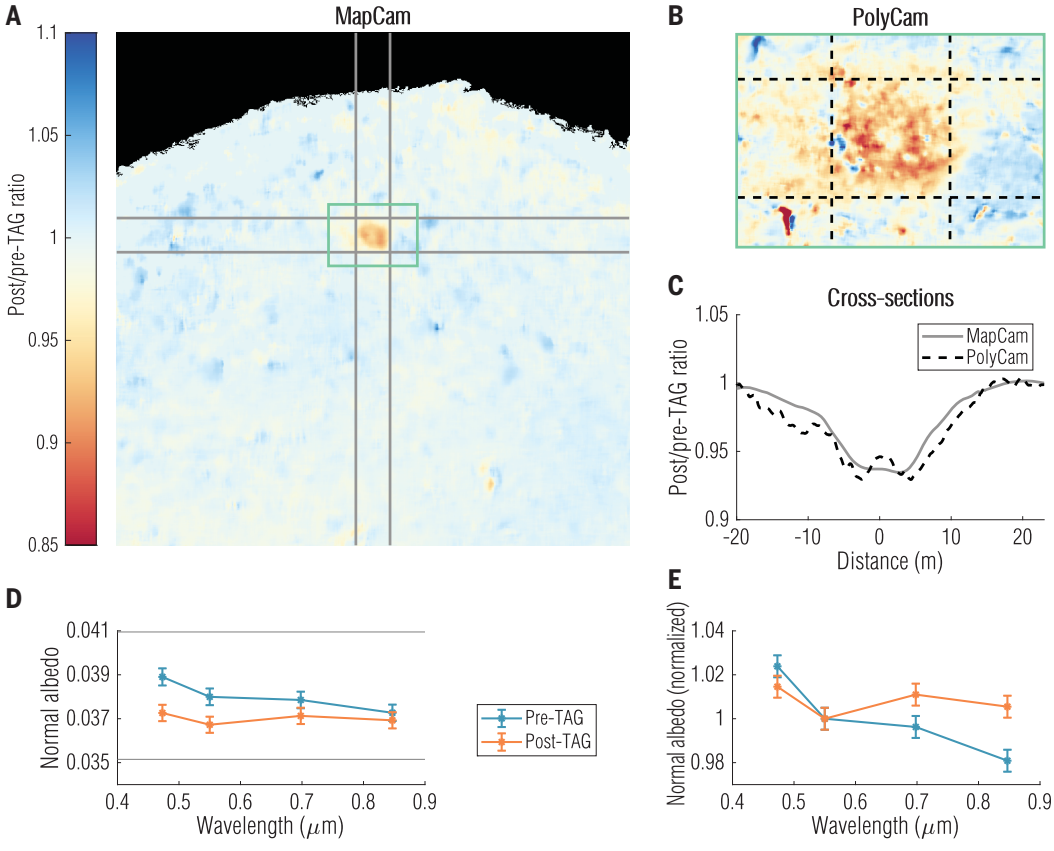


Fig. 4. Changes in surface optical properties after sampling, measured with OCAMS. (**A** and **B**) Reflectance changes in (A) MapCam and (B) PolyCam data before and after sampling (TAG). (**C**) Plot of reflectance change across Hokioi crater. (**D**) Absolute reflectance from MapCam color filters. The error bars indicate the 1% relative uncertainty between filters (*9*). The horizontal gray bars indicate the absolute radiometric uncertainty. (**E**) The same spectra normalized to 0.55 μm. The error bars indicate the relative uncertainty of ±0.25% between pre- and post-TAG imaging (*9*).

The differences between the pre- and post-sampling DTMs indicate that the elliptical shape of the TAG crater arises from two factors. First, the sampling event occurred on a 20° to 30° north-to-south slope, relative to Bennu's gravity vector (Fig. 3E). We infer that a transient crater formed when the gas release led to upslope steepening, in excess of the angle of repose, especially to the north and northwest. This interaction would have led to subsequent mass wasting, producing the elliptical shape and the rock accumulations at the crater center (Fig. 3B and fig. S8). In mass movements on Earth (*30*) and on asteroids (*31*), larger blocks can surf on the finer particles of regolith, flowing downslope, and are the last to be buried. Second, the sampling event excavated surface material along the north-south long axis of the TAG crater and along a line running along its short axis from the west-southwest to east-northeast (Fig. 3F and figs. S11 and S12). These shallow trenches of material loss correspond to the orientation of the back-away thrusters, so we conclude that they were produced when the spacecraft pulled away from the surface. We estimate that the thrusters increased the volume of the crater by ~40% (figs. S11 and S12).

Elliptical morphology and accumulation of larger rocks at the crater bottom can occur in low-velocity (<500 m s^{-1}) impact experiments on a 20° to 25° slope in Earth gravity (*32*–*34*). Surface flow due to parallel erosion of the walls of a transient crater can lead to sorting, with coarser pebbles and less dense materials floating to the top, whereas smaller fines are buried below (*31*). In such flows, large rocks tend to move to the toe of the flow and can be partially buried when they come to a stop at the center of the crater, as more flow follows behind. Depth profiles through the TAG crater indicate removal of up to 20 cm of material over a broad region in the north and north-northeast (fig. S11), along the steepest areas of the Hokioi crater wall. This portion of the wall likely readjusted through surface failure, in response to the formation of the TAG crater.

Physical properties of the subsurface

We applied crater scaling relationships (*35*–*37*) and used the known energy of the gas released to assess the physical properties of the rubble in which the TAG crater formed (table S4) (*9*). As an upper bound, we assumed that the entire 8-m (average diameter) TAG crater is the result of TAGSAM gas release. As a lower bound, we removed the possible contribution to the crater size from thruster effects, yielding a 5-m average diameter (fig. S12).

For rubble-pile asteroids, intermolecular attraction may result in cohesive forces between particles, with values approximately equal to the gravitational force (*23*, *38*). For both crater sizes considered, however, a nearly cohesionless (<0.001 Pa) granular material is required (fig. S13, G to I). Even under these conditions, the energy associated with TAGSAM gas release does not reproduce the observed crater if the bulk density of the regolith is equal to that of the global Bennu average [1190 kg m^{-3} (*5*)]. Instead, the observed crater size indicates a local bulk density of 500 to 700 kg m^{-3} (fig. S13, G to I). This range is consistent with independent results from spacecraft accelerometer data and granular mechanics models (*21*). This density, combined with the estimated excavation volume, suggests that the sampling event mobilized at least 6000 ± 650 kg, which is substantially more than predicted (*39*).

Particle size differences

We counted particles in PolyCam images of the sample site (fig. S14) taken during the final flyby to determine the PSFD of the newly

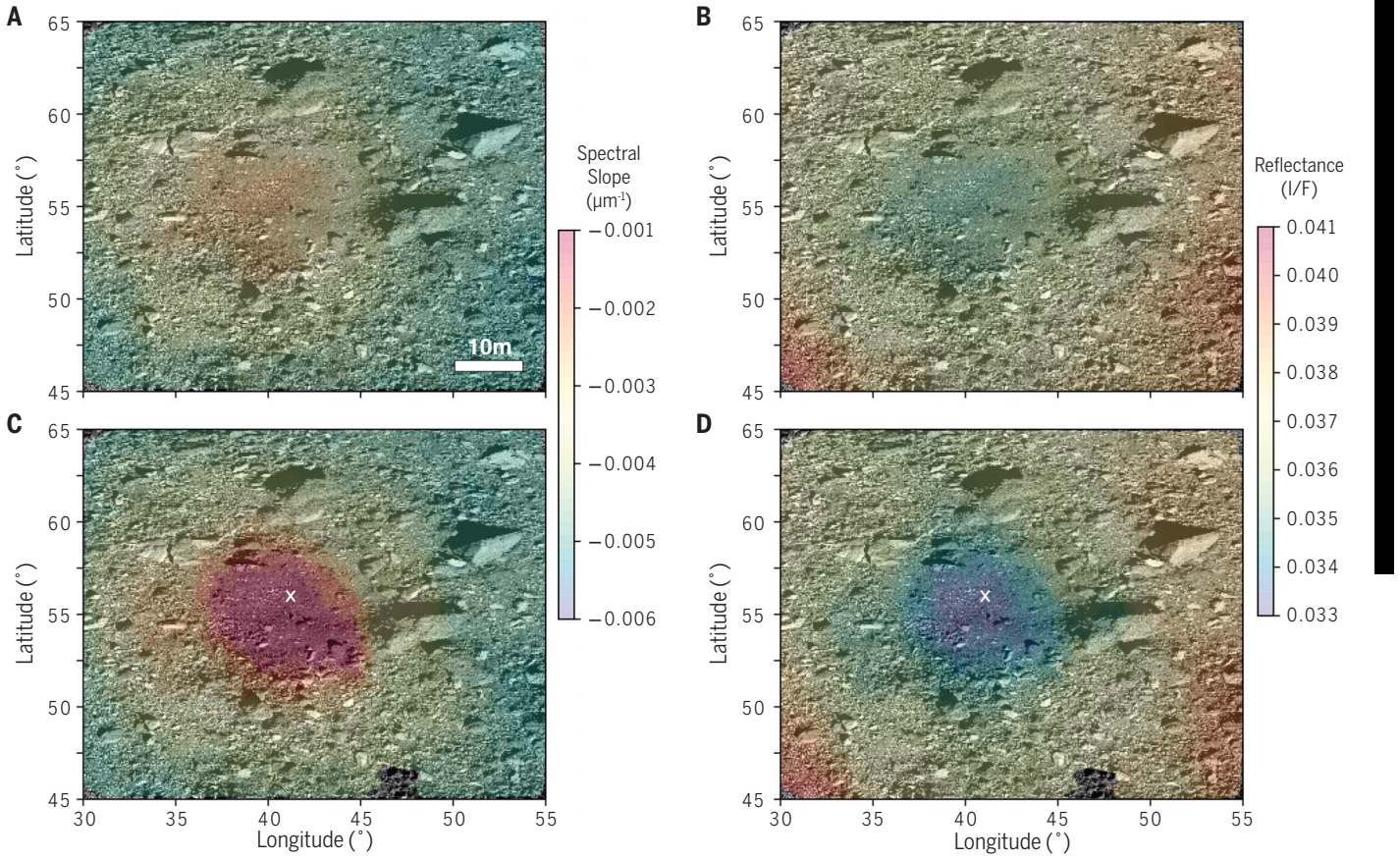


Fig. 5. Changes in surface spectral properties measured with OVIRS. Data collected with OVIRS (color) are overlain on a pre-sampling PolyCam image of the sample site. (**A**) Spectral slope from 0.5 to 1 μm before sampling. (**B**) Reflectance at 0.55 μm before sampling. (**C** and **D**) Same as (A) and (B), but after sampling. The sampling location is indicated with the white cross.

exposed subsurface (fig. S9 and table S1) (*9*). Before sampling, particle counts from PolyCam images of Nightingale, obtained at the same imaging conditions as the final flyby, yielded a PSFD power-law index of −1.9 ± 0.2 (*11*). The final flyby data have a power-law index of −2.5 ± 0.1 (fig. S9). The sampling event thus decreased the exponent in the power-law distribution of visible particles. The change in PSFD is likely a combination of particle fragmentation from TAG combined with a subsurface reservoir of fine particles (Fig. 2, A and B).

Optical and spectral differences

Post-sampling MapCam and PolyCam data (figs. S15 and S16) (*9*) show that Nightingale is darker overall, with normal albedo decreased by 5% (Fig. 4, A to C). However, a greater amount of high-reflectance material [relative to the generally dark surface (*5*, *14*, *28*)] is visible near the contact point after sampling (Fig. 3, A and B). MapCam color data show that the Nightingale surface became spectrally redder in the wavelength range 0.55 to 0.85 μm and an absorption band near 0.55 μm, which is indicative of magnetite, is deeper than before sampling (Fig. 4, D and E, and table S5) (*5*).

The post-sampling surface is also spectrally redder and darker in OVIRS spectra (Fig. 5), compared with data collected before sampling (*6*). The visible–near-infrared spectral slope (0.5 to 1 μm) at the sampling site increased by 58% (figs. S17 and S18 and table S5). Reflectance at 0.55 μm decreased by ~5% relative to previous observations, after accounting for reduced optical throughput (*9*). These changes do not extend past the edge of Hokioi crater (fig. S18 and table S5). The OVIRS detector was too warm during the post-sampling observations to determine changes at longer wavelengths. Comparison of pre-sampling (*13*) and post-sampling OTES spectra shows no detectable change in thermal properties of the surface at Nightingale (figs. S19 to S21) (*9*).

Previous work on color variation across Bennu's surface proposed that the most freshly exposed surfaces are among the reddest and darkest (*14*); our results support this interpretation. We attribute the spectral changes to the exposure of fresh, organic-rich material (supplementary text).

Properties of the collected sample

In-flight inspection of the sample

SamCam took images of TAGSAM 2 days after sampling (22 October 2020). SamCam and the TAGCAMS imager StowCam acquired further images 8 days after sampling (28 October 2020), just before TAGSAM was stowed in its protective return capsule (*9*). These images showed that the TAGSAM collection chamber was packed with sample (fig. S22), with the average brightness of the chamber reduced to 1.5% of its pre-sampling value. The images also showed 23 particles, up to 4 mm in diameter, clinging to the TAGSAM contact pads (fig. S23 and table S6). The pattern of specular and diffuse reflection on each pad across different illumination conditions indicates the presence of unresolved particles. More than 260 submillimeter particles adhered directly to other parts of TAGSAM (fig. S23 and table S6). We anticipate that submillimeter dust mobilized by the sampling event (Fig. 2, A and B) is also trapped in the pads.

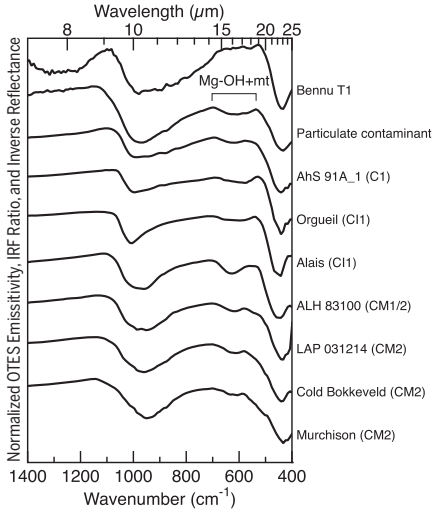


Fig. 6. Spectroscopic comparison of Bennu to carbonaceous chondrite meteorites. From top to bottom, the lines are an example Bennu surface spectrum acquired by OTES before sampling [type T1 (*13*)]; the spectrum of the contaminating particulates on the instrument mirror after sampling; and spectra of aqueously altered CI and CM chondrites (*13*), which are proposed Bennu-analog meteorites. Spectra are offset vertically for clarity; the Bennu T1 spectrum is scaled by 300% for comparison of spectral features. The small Mg-OH plus magnetite (mt) absorption from ~700 to 535 cm^{-1} (~14.3 to 18.7 μm) in the particulate spectrum was not detectable in Bennu surface observations but is common among the CIs and CMs. IRF, instrument response function.

These images also showed that TAGSAM was overflowing and losing particles (fig. S24). Several pebbles with long axes up to 3 cm were visibly wedged in the chamber mouth, propping open the mylar flap that was intended to prevent material from escaping (fig. S22).

To characterize the sample loss, we tracked 1804 individual particles leaving TAGSAM in the SamCam verification images taken 2 days after sampling (*9*). During this verification sequence, the TAGSAM wrist motor performed nine different articulations, each imaged by SamCam with multiple exposure times. Each tracked particle was observed in at least three images, allowing their trajectories to be estimated and traced back to TAGSAM, by using methods previously developed to track particles ejected from Bennu's surface (*26*, *27*, *40*). The median estimated range from the camera to the particles at the time of their first observation was 2.4 m, and the median estimated particle velocity was 0.8 cm s^{-1} (fig. S25). We observed the largest number of particles (560, or 31% of all escaping particles observed) leaving just after the first movement of the wrist (fig. S24A). Fewer escaping particles were observed with each successive wrist movement (fig. S24, B to H), until the wrist reversed direction, at which point the number of escaping particles increased (fig. S24I).

We estimated the size and mass of each tracked particle (fig. S25) (*9*). The escaping particles had a flake-like shape (fig. S26), which is consistent with particles naturally ejected from Bennu (*27*, *41*). Their long dimensions ranged from 0.004 to 23 mm, with a mean of 0.92 mm. We estimate that at least 55 g of material left TAGSAM, with a mean particle mass of 0.03 g and an average mass loss rate of 2.2 g min^{-1} throughout the imaging sequence. The tracked particles are only a fraction of the total particles in the images, which in turn are likely only a fraction of the total particles that escaped. Thus, these results represent a lower bound on the sample mass lost.

Sample mass from momentum transfer analysis

To estimate the mass of the collected sample, the spacecraft was programmed to hold an inertially fixed attitude while the TAGSAM arm moved to determine the exchange of momentum between TAGSAM and the reaction wheels (*42*). This activity was performed three times: before sampling (zero mass), 2 days after sampling (concurrent with SamCam verification imaging), and 8 days after sampling (concurrent with SamCam and StowCam verification imaging). Analysis of the momentum transfer (*42*) indicates that at 2 days after sampling, TAGSAM contained 317 ± 101 g of sample. This mass is consistent with a prediction of 252 to 575 g (*8*) based on empirical formulations of TAGSAM performance under the observed sampling conditions. By the next measurement (8 days after sampling), just before sample stowage, the same technique was used to measure a sample mass of 250 ± 101 g (*42*), indicating that TAGSAM lost 67 g in the interim.

Size-frequency distribution of particles on TAGSAM

We measured the diameters of particles on top of TAGSAM in SamCam images during back-away. Even though these particles were removed before sample stowing, we obtained their PSFD as an approximation for that of the captured sample (fig. S9A and table S1) (*9*). The obtained PSFD follows a power law with an average index of −2.2 ± 0.3. This value is consistent with the power-law index of −2.3 ± 0.3

measured at Nightingale before sampling (*11*). However, given the observed friability of Bennu particles, the PSFD of the sample after experiencing atmospheric entry forces upon delivery to Earth in 2023 might not resemble that of either Nightingale or the initially collected sample.

Composition of particulates on instrument optics

The mobilization of a dust cloud and debris plume resulted in the accumulation of particulates on the optics of the OSIRIS-REx science instruments (table S7) (*9*). After sample collection, OTES was found to have experienced a ~15% decrease in signal, of which only about ~1% can be attributed to fluctuations in the measured instrument temperature. We used this optical contamination to spectrally characterize the particulates (Fig. 6) by taking the ratio of OTES observations of space before and after sampling (*9*). Because the instrument, and thus the particulates, are hotter than the background space, the contaminant spectrum is analogous to an emission spectrum (*43*).

This spectrum resembles that of Bennu's surface (*12*, *13*), with features attributed to stretching and bending modes in phyllosilicate minerals (Fig. 6). Absorption bands due to magnetite are deeper than on Bennu's surface. In addition, we observed a compound absorption band centered near 605 cm^{-1} (~16.5 μm), which we attribute to an Mg-OH stretching mode in hydrated Mg-rich phyllosilicates, plus a stretching mode in magnetite (*44*, *45*).

Mg-rich phyllosilicates are diagnostic of high degrees of aqueous alteration (*1*)—that is, secondary mineralization owing to rock-fluid interactions that took place early in Solar System history, on the precursor bodies of carbonaceous asteroids. Previous analyses of OTES data posited that Bennu's phyllosilicates were Mg-rich (*12*), but we could not confirm this because the 605 cm^{-1} compound absorption was not detectable and the silicate stretching band was distorted by thin dust deposits (*13*). Our post-sampling particulate spectrum substantiates the previously published interpretation (*12*) of Bennu as analogous to the most aqueously altered, chemically primitive (*46*) carbonaceous meteorites: the CI (Ivuna-type) and CM (Mighei-type) carbonaceous chondrites.

Conclusion

TAGSAM collected a much greater sample than the mission's requirement of 60 g (*3*, *4*), which we attribute to the soft, noncohesive, low-bulk-density surface and upper subsurface (top ~50 cm) of Bennu. Although large particles (a few centimeters) wedged in the TAGSAM flap led to the loss of some collected sample, hundreds of grams were stowed for return to Earth (*42*). Our observations of dust mobilization and surface changes suggest that the sample may have different textural properties than the surface and should include material that has undergone different degrees of space weathering (*14*).

REFERENCES AND NOTES

1. D. S. Lauretta, H. Y. McSween, Eds., *Meteorites and the Early Solar System II* (Univ. Arizona Press, 2006).
2. C. Chyba, C. Sagan, *Nature* **355**, 125–132 (1992).
3. D. S. Lauretta *et al*., *Space Sci. Rev.* **212**, 925–984 (2017).
4. D. S. Lauretta, H. L. Enos, A. T. Polit, H. L. Roper, C. W. V. Wolner, in *Sample Return Missions*, A. Longobardo, Ed. (Elsevier, 2021), pp. 163–194.
5. D. S. Lauretta *et al*., *Nature* **568**, 55–60 (2019).
6. A. A. Simon *et al*., *Science* **370**, eabc3522 (2020).
7. D. J. Scheeres *et al*., *Sci. Adv.* **6**, eabc3350 (2020).
8. K. J. Walsh *et al*., *Space Sci. Rev.* **218**, 20 (2022).
9. Materials and methods are available as supplementary materials.
10. K. Berry *et al*., 43rd Annual AAS Guidance, Navigation and Control Conference, Breckenridge, CO, paper AAS 20-088 (2020).
11. K. N. Burke *et al*., *Remote Sens.* **13**, 1315 (2021).
12. V. E. Hamilton *et al*., *Nat. Astron.* **3**, 332–340 (2019).
13. V. E. Hamilton *et al*., *Astron. Astrophys.* **650**, A120 (2021).
14. D. N. DellaGiustina *et al*., *Science* **370**, eabc3660 (2020).
15. A. A. Simon *et al*., *Astron. Astrophys.* **644**, A148 (2020).
16. H. H. Kaplan *et al*., *Astron. Astrophys.* **653**, L1 (2021).
17. H. H. Kaplan *et al*., *Science* **370**, eabc3557 (2020).
18. B. Rozitis *et al*., *Sci. Adv.* **6**, eabc3699 (2020).
19. B. Rozitis *et al*., *J. Geophys. Res. Planets* **125**, e2019JE006323 (2020).
20. M. Delbó, P. Michel, *Astrophys. J. Lett.* **728**, L42 (2011).
21. K. J. Walsh *et al*., *Sci. Adv.* 10.1126/sciadv.abm6229 (2022).
22. R. Ballouz, "Numerical simulations of granular physics in the solar system," thesis, University of Maryland, College Park, MD, 2017.
23. R. Ballouz *et al*., *Mon. Not. R. Astron. Soc.* **507**, 5087–5105 (2021).
24. N. Murdoch *et al*., *Mon. Not. R. Astron. Soc.* **503**, 3460–3471 (2021).
25. R.-L. Ballouz *et al*., *Nature* **587**, 205–209 (2020).
26. D. S. Lauretta *et al*., *Science* **366**, eaay3544 (2019).
27. C. W. Hergenrother *et al*., *J. Geophys. Res. Planets* **125**, e2020 (2020).
28. D. N. DellaGiustina *et al*., *Nat. Astron.* **3**, 341–351 (2019).
29. E. B. Bierhaus *et al*., *Nat. Geosci.* **15**, 440–446 (2022).
30. R. M. Iverson, M. E. Reid, R. G. LaHusen, *Annu. Rev. Earth Planet. Sci.* **25**, 85–138 (1997).
31. V. Perera, A. P. Jackson, E. Asphaug, R.-L. Ballouz, *Icarus* **278**, 194–203 (2016).
32. J. Aschauer, T. Kenkmann, *Icarus* **290**, 89–95 (2017).
33. K. Hayashi, I. Sumita, *Icarus* **291**, 160–175 (2017).
34. R. T. Daly *et al*., in *Proceedings of the 15th Hypervelocity Impact Symposium, Destin, FL, USA* (American Society of Mechanical Engineers, 2019), pp. 237–244.
35. K. A. Holsapple, *Annu. Rev. Earth Planet. Sci.* **21**, 333–373 (1993).
36. K. R. Housen, K. A. Holsapple, *Icarus* **211**, 856–875 (2011).
37. K. R. Housen, R. M. Schmidt, K. A. Holsapple, *J. Geophys. Res.* **88** (B3), 2485–2499 (1983).
38. D. C. Richardson, P. Michel, K. J. Walsh, K. W. Flynn, *Planet. Space Sci.* **57**, 183–192 (2009).
39. E. B. Bierhaus *et al*., *Icarus* **355**, 114142 (2021).
40. J. Y. Pelgrift *et al*., *Earth Space Sci.* **7**, e2019EA000938 (2020).
41. S. R. Chesley *et al*., *J. Geophys. Res. Planets* **125**, e2019JE006363 (2020).
42. H. Ma, M. Skeen, R. Olds, B. Miller, D. S. Lauretta, arXiv:2109.05561 [astro-ph.EP] (2021).
43. S. W. Ruff *et al*., *J. Geophys. Res.* **111**, E12S18 (2006).
44. R. D. Hanna *et al*., *Icarus* **346**, 113760 (2020).
45. M. Ishii, M. Nakahira, T. Yamanaka, *Solid State Commun.* **11**, 209–212 (1972).
46. N. Grevesse, M. Asplund, J. Sauval, *Space Sci. Rev.* **130**, 105–114 (2007).
47. B. Rizk, C. Drouet d'Aubigny, D. Golish, D. N. DellaGiustina, D. S. Lauretta, NASA Planetary Data System, urn:nasa:pds:orex.ocams, vversion 10.0 (2019); https://arcnav.psi.edu/urn:nasa:pds:context:instrument:ocams.orex.
48. C. Bos, Jackman, D.S. Lauretta, NASA Planetary Data System, urn:nasa:pds:orex.tagcams, version 10.0 (2019); https://arcnav.psi.edu/urn:nasa:pds:context:instrument:tagcams.orex.
49. P. Christensen, V. Hamilton, S. Anwar, G. Mehall, D. S. Lauretta, NASA Planetary Data System, urn:nasa:pds:orex.otes, version 10.0 (2019); https://arcnav.psi.edu/urn:nasa:pds:context:instrument:otes.orex.
50. D. C. Reuter, A. A. Simon, A. Lunsford, D. S. Lauretta, NASA Planetary Data System, urn:nasa:pds:orex.ovirs, version 10.0 (2019); https://arcnav.psi.edu/urn:nasa:pds:context:instrument:ovirs.orex.
51. M. Daly, O. Barnouin, R. Espiritu, D. Lauretta, NASA Planetary Data System, urn:nasa:pds:orex.ola, version 3.0 (2019); https://arcnav.psi.edu/urn:nasa:pds:context:instrument:ola.orex.

ACKNOWLEDGMENTS

The OSIRIS-REx sample collection event was the culmination of years of effort by hundreds of team members in the face of substantial challenges. We are grateful to all who contributed. **Funding:** D.S.L., C.W.V.W., C.D.A., D.S.N., J.Y.P., E.M.S., E.J.L.-C., L.K.M., C.W.H., K.J.W., R.L.B., O.S.B., D.N.D., K.J.B., T.B., C.A.B., D.R.G., C.Y.D.d'A., B.Ri., B.J.B., J.B.G., M.A.R., R.B., M.C.M., H.C., A.A., J.N., V.E.H., P.R.C., A.A.S., B.E.C., H.H.K., D.C.R., H.L.E., A.T.P., M.M.W., S.H.S., K.G., A.J.L., L.P.K., R.O., H.M., B.D.M., M.A.S., S.F.K., E.C.A.C., A.J.R., J.P.E., and H.C.C. were supported by NASA under contract NNM10AA11C issued through the New Frontiers Program. M.P. was supported by the Italian Space Agency (ASI) under the ASI-INAF agreement 2017-37-H.0. S.S. was supported by the International Network of Planetary Sciences grant from the Japanese Society for Promotion of Science (JSPS) Core-to-Core program. B.Ro. acknowledges funding support from the UK Science and Technology Facilities Council (STFC). M.G.D. and J.A.S. were supported by the Canadian Space Agency. **Author contributions:** D.S.L. led the investigation and compiled, integrated, and interpreted the results, supported by C.W.V.W.; C.D.A. led the investigation of escaping particles from TAGSAM, supported by D.S.N., J.Y.P., E.M.S., E.J.L.-C., and L.K.M.; C.W.H. analyzed the shapes of particles escaping from TAGSAM and supported their photometric analysis. K.J.W. led the evaluation of regolith properties, supported by R.L.B.; O.S.B. led the altimetry investigation of the TAG crater, supported by M.G.D. and J.A.S.; M.G.D. is the instrument scientist for OLA. D.N.D. led the image processing of OCAMS data, supported by K.J.B., T.B., C.A.B., D.R.G., C.Y.D.d'A., and B.Ri.; C.Y.D.d'A. and B.Ri. are the OCAMS instrument scientists. E.B.B. is the TAGSAM lead scientist and analyzed the thruster contribution to surface modification. B.J.B. is the TAGCAMS instrument scientist and led the image processing of TAGCAMS data, supported by J.B.G. and M.A.R.; R.B. and M.C.M. are the project managers for the OSIRIS-REx mission. M.P. led the particle size frequency analysis of the final flyby data, supported by H.C., A.A., and J.N.; S.S. performed the particle size frequency analysis of the TAGSAM head and the debris plume, supported by Y.C., T.M., N.S., and K.Y.; V.E.H. led the spectral analysis of OTES data, supported by P.R.C.; P.R.C. and V.E.H. are the OTES instrument scientists. A.A.S. led the spectral analysis of OVIRS data supported by B.E.C., H.H.K., and D.C.R.; D.C.R. and A.A.S. are the OVIRS instrument scientists. H.L.E. led the science operations team, supported by A.T.P., M.M.W., and S.H.S.; K.G. led the flight dynamics and navigation team, supported by C.D.A., D.S.N., J.Y.P., E.M.S., E.J.L.-C., A.J.L., and L.K.M.; L.P.K. led the analysis of the contact pads, supported by B.Ri.; R.O. led the analysis of spacecraft momentum buildup, supported by H.M., B.D.M., and M.A.S.; S.F.K. led the spacecraft operations team. E.C.A.C. was the TAG lead system engineer. B.Ro. led the thermal analysis, supported by A.J.R. and J.P.E.; H.C.C.Jr. is the mission sample scientist and supported interpretation of the collected sample. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Data from the sampling event and/or the final flyby are available the Planetary Data System at https://sbn.psi.edu/pds/resource/orex in the datasets devoted to OCAMS (*47*), TAGCAMS (*48*), OTES (*49*), OVIRS (*50*), and OLA (*51*). The IDs of the images, spectra, and scans we used are provided in data S1. Our pre- and post-sampling DTMs are provided as data S2 and S3 or can be accessed in the Small Body Mapping Tool (SBMT; https://sbmt.jhuapl.edu) by following the instructions in data S1.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm1018
Materials and Methods
Supplementary Text
Figs. S1 to S26
Tables S1 to S7
References (*52*–*90*)
Movies S1 to S6
Data S1 to S3

Submitted 28 August 2021; accepted 10 June 2022
Published online 7 July 2022
10.1126/science.abm1018

HEART DISEASE

Hematopoietic loss of Y chromosome leads to cardiac fibrosis and heart failure mortality

Soichi Sano[1,2]*, Keita Horitani[1], Hayato Ogawa[1]†, Jonatan Halvardson[3], Nicholas W. Chavkin[1,4], Ying Wang[1]‡, Miho Sano[1]§, Jonas Mattisson[3], Atsushi Hata[5], Marcus Danielsson[3], Emiri Miura-Yura[1], Ammar Zaghlool[3], Megan A. Evans[1], Tove Fall[6], Henry N. De Hoyos[1], Johan Sundström[7]¶, Yoshimitsu Yura[1]†, Anupreet Kour[1], Yohei Arai[1], Mark C. Thel[1], Yuka Arai[1], Josyf C. Mychaleckyj[8], Karen K. Hirschi[4], Lars A. Forsberg[3,9]*, Kenneth Walsh[1]*

Hematopoietic mosaic loss of Y chromosome (mLOY) is associated with increased risk of mortality and age-related diseases in men, but the causal and mechanistic relationships have yet to be established. Here, we show that male mice reconstituted with bone marrow cells lacking the Y chromosome display increased mortality and age-related profibrotic pathologies including reduced cardiac function. Cardiac macrophages lacking the Y chromosome exhibited polarization toward a more fibrotic phenotype, and treatment with a transforming growth factor β1–neutralizing antibody ameliorated cardiac dysfunction in mLOY mice. A prospective study revealed that mLOY in blood is associated with an increased risk for cardiovascular disease and heart failure–associated mortality. Together, these results indicate that hematopoietic mLOY causally contributes to fibrosis, cardiac dysfunction, and mortality in men.

The human male-specific Y chromosome is relatively small in size and contains a limited number of genes that regulate sex determination and spermatogenesis (*1*). Beyond sex determination, there is a paucity of information about the biological role of the Y chromosome, partly because of challenges in determining genetic variation caused by the inter- and intrachromosomal repeat sequences. However, insights into the physiological role of the Y chromosome can be aided by studies that address the mosaic loss of chromosome Y (mLOY) in the blood, a condition in which a fraction of hematopoietic cells display a loss of the Y chromosome. This phenomenon is the most prevalent post-zygotic mutation in leukocytes (*2*). The frequency of hematopoietic mLOY increases with age and smoking status (*3*, *4*) and is associated with the condition of clonal hematopoiesis of indeterminate potential (CHIP) (*5*, *6*). Although the technology to assess mLOY is evolving, it has recently been reported that mLOY is detectable in 40% of 70-year-old males and 57% of 93-year-old males (*4*, *7*). Loss of the Y chromosome is prevalent in hematologic malignancies and may be a factor in the prognosis of these diseases. Whereas most men with mLOY never progress to a hematologic cancer, epidemiological studies have shown that mLOY in blood is associated with shorter life span (*2*, *8*) and increased incidence of various age-associated diseases, including solid tumors and Alzheimer's disease (*2*, *9*). Furthermore, mLOY has previously been associated with secondary major cardiovascular events in atherosclerotic patients after carotid endarterectomy (*10*) and with prior heart attack and stroke self-reported at baseline in the UK Biobank study (*8*). It has been reported that mLOY is in part a manifestation of inherited genomic instability and a marker of biological aging (*4*). However, because of the descriptive nature of epidemiological studies, it is unknown whether mLOY plays a causal role in disease development. Here, we modeled hematopoietic mLOY in mice and examined its role in fibrosis and cardiac dysfunction. Together with these results, prospective analyses of UK Biobank data suggest that mLOY in leukocytes directly contributes to heart failure.

[1]Hematovascular Biology Center, Robert M. Berne Cardiovascular Research Center, University of Virginia School of Medicine, Charlottesville, VA 22908, USA. [2]Department of Cardiovascular Medicine, Osaka Metropolitan University Graduate School of Medicine, Osaka 545-8585, Japan. [3]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, 75108 Uppsala, Sweden. [4]Department of Cell Biology, University of Virginia School of Medicine, Charlottesville, VA 22908, USA. [5]Chiba University Graduate School of Medicine, Department of General Thoracic Surgery, Chiba 260-8670, Japan. [6]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, 75185 Uppsala, Sweden. [7]Department of Medical Sciences, Uppsala University, Sweden, and Uppsala Clinical Research Center, 78185 Uppsala, Sweden. [8]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA. [9]The Beijer Laboratory, Uppsala University, 75185 Uppsala, Sweden.
*Corresponding author. Email: sano.soichi@omu.ac.jp (S.S.); lars.forsberg@igp.uu.se (L.A.F.); kw9ar@virginia.edu (K.W.)
†Present address: Department of Cardiology, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan.
‡Present address: Department of Cardiology, Xinqiao Hospital, Army Medical University, Chongqing 400037, China.
§Present address: Department of Obstetrics and Gynecology, Osaka Metropolitan University Graduate School of Medicine, Osaka 545-8585, Japan.
¶Present address: The George Institute for Global Health, University of New South Wales, Newtown, New South Wales 2042, Australia.

Results

Age-related pathologies are accelerated in mice with mLOY in blood

We constructed a mouse model of hematopoietic mLOY by using CRISPR-Cas9 gene editing to target repeat DNA sequences that are specific to the centromere of the Y chromosome. Guide RNAs (gRNAs) and a tomato red fluorescent protein (tRFP) marker were delivered to lineage-negative bone marrow cells through the lentivirus vector before transplantation into lethally irradiated, wild-type male mice (Fig. 1A). To decouple the *Cas9* gene from the lentivirus vector and to maximize gene-editing efficiency, donor bone marrow cells were isolated from ROSA26-Cas9 knock-in mice that express the Cas9 endonuclease systemically (*11*). Two LOY-gRNAs, LOY-gRNA1 and LOY-gRNA2, targeting different repeat sequences within the centromere were evaluated for the efficiency of Y chromosome ablation by fluorescent in situ hybridization (FISH) analysis of X and Y chromosomes in tRFP$^+$ blood cells collected from mice reconstituted with bone marrow cells. As a control, we used a lentivirus encoding a gRNA designed to not target any region of the genome in the ROSA26-Cas9 knock-in donor cells. Y chromosome ablation efficiency was ~95 and 80% for LOY-gRNA1 and LOY-gRNA2, respectively (Fig. 1, B and C). Thus, unless otherwise indicated, the mouse model of mLOY used a lentivirus vector that expresses the LOY-gRNA1 transcript and the tRFP$^+$ marker protein to assess cell transduction. Y chromosome ablation in vivo was validated by karyotype analysis using lineage-negative bone marrow cells isolated from mLOY and control mice that were immortalized by lentivirus-mediated HoxB8 overexpression (Fig. 1D). Consistent with the ablation of the male sex chromosome, the Y chromosome–encoded transcripts *Kdm5d*, *Uty*, *Eif2s3y*, and *Ddx3y* were not detectable in the circulating tRFP$^+$ leukocytes of the mLOY mice (Fig. 1E). Because of the inefficiency of bone marrow progenitor cell transduction by the lentivirus vector, blood chimerism, defined as a percentage of tRFP$^+$ white blood cells, ranged from 49 to 81% in various experiments (mean = 64.9 ± 4.0%), which is consistent with the levels of mLOY in men that have been associated with various disease processes (*2*, *9*). These levels of chimerism were maintained for 12 months (fig. S1A). Focusing on each immune cell population, chimerism was higher in myeloid cells compared with B and T cells, consistent with observations of mLOY in men (*12*) (fig. S1B).

The phenotypic consequences of mLOY were evaluated in aging male mice. No obvious hematological abnormalities were observed in mLOY or control mice during the follow-up period (fig. S1C). However, mice with the mLOY condition displayed shorter life spans compared with control mice (Fig. 1F). Serial echocardiographic analyses revealed the development of an accelerated age-associated cardiomyopathy in the mLOY mice, with greater

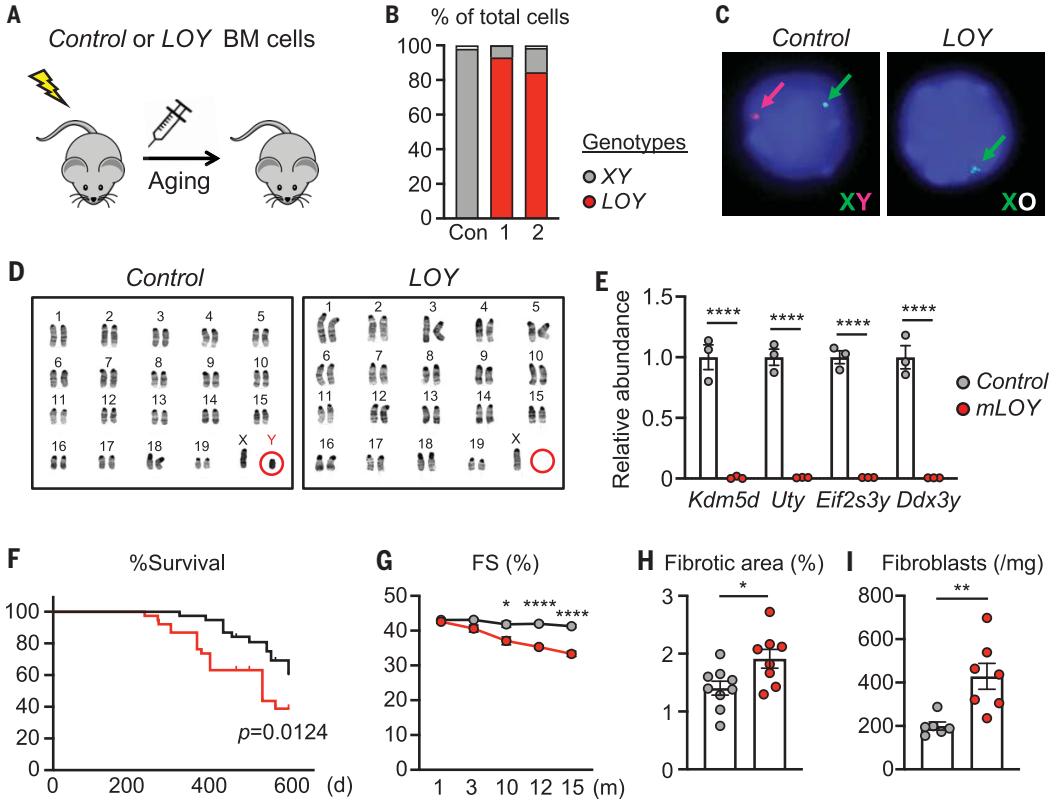
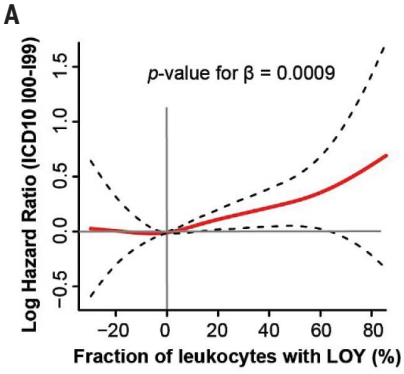


Fig. 1. Y chromosome deficiency in hematopoietic cells shortens life span and accelerates age-related cardiac dysfunction. (**A**) Schematic of this study. Lethally irradiated male C57BL6/J mice were reconstituted with hematopoietic stem cells transduced with lentivirus encoding Y chromosome targeting gRNA (LOY-gRNA) or control gRNA and designated as mLOY and control mice, respectively. Phenotypic differences between mLOY and control mice during the natural aging process were analyzed. (**B**) Efficiency of Y chromosome ablation analyzed by FISH. tRFP^{+} peripheral blood cells were collected from mice reconstituted with bone marrow cells transduced with lentivirus encoding a control gRNA or either of two different LOY-gRNAs: LOY-gRNA1 or LOY-gRNA2. The percentages of Y chromosome–deficient (LOY) cells and Y chromosome–sufficient cells (XY) in total blood cells are shown. Approximately 200 cells were analyzed for each condition. Unless otherwise indicated, all subsequent studies were performed with hematopoietic stem cells transduced with LOY-gRNA1. (**C**) Representative images of FISH analysis of peripheral blood cells collected from mLOY and control mice. Green and red fluorescence indicate the X and Y chromosomes, respectively. (**D**) Karyotype analysis of LOY and control cells. Hematopoietic stem cells collected from mLOY and control mice were immortalized by lentivirus-mediated HoxB8 transduction and subjected to karyotype analysis. (**E**) mRNA expression of genes on the Y chromosome in RFP^{+} peripheral blood leukocytes in mLOY and control mice (n = 3 per group). (**F**) Kaplan-Meier survival curve for mLOY and control mice after BMT. The x axis indicates time after BMT (day) (control, n = 37; mLOY, n = 38). (**G**) Sequential echocardiographic analysis of mLOY and control mice after BMT at the indicated time points (month) (n = 8 to 10 per group). (**H**) Quantitative analysis of fibrotic area in heart section at 15 months after BMT (n = 8 to 9 per group). (**I**) Flow cytometric analysis of fibroblast counts in heart tissue at 15 months after BMT (n = 6 to 7 per group). The absolute numbers of cells were normalized by tissue weight. Dots in all panels represent individual samples. Data are shown as mean ± SEM. Statistical analyses were performed using unpaired Student's t test [(E) and (H)], Student's t test with Welch's correction (I), log-rank test (F), and two-way repeated-measures ANOVA with Sidak's multiple-comparisons tests (G). FS, fractional shortening; RFP; red fluorescent protein, Con; control. *P < 0.05, **P < 0.01, ****P < 0.0001.

cardiac dysfunction detected in older mice (Fig. 1G and fig. S2A). Consistent with the development of an age-associated cardiomyopathy, mLOY mice displayed a small increase in heart mass at the termination of the experiment, but there were no differences in body weights between the mLOY and control genotypes over the course of the experiment (fig. S2B). The mLOY condition led to an increase in myocardial fibrotic area (Fig. 1H), as determined by quantitative analysis with picrosirius red staining of cardiac sections (fig. S2C), and an increase in the quantity of myocardial MEF-SK4^{+} fibroblasts, as assessed by flow cytometry (Fig. 1I). Consistent with this fibrotic response, the mLOY condition also led to increased left ventricular filling pressure, which is indicative of diastolic dysfunction (fig. S2D). These cardiac changes were observed despite modest reductions in blood pressure in the mLOY mice (fig. S2E) and no changes in the serum concentrations of renin 1 and angiotensin II (fig. S2F). Although further analyses of how mLOY affects the renin-angiotensin system are warranted, these experimental results are consistent with observations of diminished self-reported hypertension in cardiovascular disease patients with mLOY (*10*). At the 15-month time point after bone marrow transplantation (BMT), mLOY mice also showed accelerated fibrotic response in the lung interstitium (fig. S3A), and greater pulmonary fibrosis could also be observed in young mice after the intratracheal administration of bleomycin (fig. S3B). Histological analysis of the kidneys also revealed a greater degree of fibrosis in the mLOY mice compared with control mice at 15 months after BMT (fig. S3C). Finally, assessments of cognitive function revealed that aging mLOY mice (15 months after BMT) had short-term working memory deficits in the Y-maze and novel object recognition tests, but these phenotypes were not observed in young mLOY mice (2 months after BMT) (fig. S4). Collectively, these results indicate that mLOY mice can recapitulate aspects of the mLOY phenotype observed in men and suggest that accelerated tissue fibrosis could be a mechanistic feature of this condition. Because a potential relationship between mLOY and cardiac dysfunction has not been reported previously, further investigations focused on the role of mLOY in heart failure are warranted.

mLOY is associated with death from cardiovascular disease and heart failure in men

mLOY has been associated with preexisting cardiovascular disease (*8*, *10*). However, a prospective analysis of mLOY with mortality from cardiovascular disease with substantial follow-up time is lacking. Thus, we investigated male survival data extracted from the UK Biobank in November 2020 (median follow-up time, 11.5 years) and conducted three different types of analysis to evaluate whether mLOY in leukocytes was associated with risk for death



B

Cause of death	ICD10	n	HR (95% CI)	Adj. *p*-val.
All causes	A00-U99	15743	1.41 (1.20-1.65)	<0.0001
Diseases of the circulatory system	I00-I99	6222	1.31 (1.02-1.70)	0.0382
Hypertensive heart disease	I11	294	3.48 (1.54-7.89)	0.0028
Heart failure	I50	997	1.76 (1.01-3.05)	0.0446
Congestive heart failure	I50.0	381	2.42 (1.14-5.15)	0.0214
Aortic aneurysm and dissection	I71	227	2.76 (1.21-6.29)	0.0154

Fig. 2. Mosaic LOY in leukocytes is associated with all-cause mortality and death caused by cardiovascular diseases in men. (**A**) Results from the primary multivariable-adjusted Cox proportional hazards regression for mLOY and mortality from diseases of the circulatory system (ICD10 codes I00 to I99) modeled with a penalized spline approach and mLOY as a continuous variable. The solid red line represents the strength of association over the spectra of LOY mosaicism, and the dotted black lines denote the 95% confidence limits of the model. A linear regression analysis supported a positive slope of the association (*P* value for β = 0.0009). The analyzed dataset was extracted from the UK Biobank with an average follow-up time of 11.5 years, and the model was adjusted for confounding effects from age, smoking, alcohol consumption, body mass index, ancestry, blood pressure, total cholesterol, and diabetes. (**B**) Results from exploratory analyses using analogous multivariable adjusted Cox models with mLOY as a binary variable (at a 40% threshold). Shown are results from analyses of all-cause mortality (ICD10 codes A00 to U99), all diseases of the circulatory system (ICD10 codes I00 to I99), as well as specific cardiovascular causes of death with at least 200 events and significant association with mLOY at the 0.05 α level. *n*, number of events.

caused by cardiovascular diseases. The percentage of blood cells lacking chromosome Y in each male participant at baseline was estimated using intensity data from single-nucleotide polymorphism array experiments (fig. S5 and table S1). In the first analysis, a continuous estimate of mLOY was used as an explanatory variable in a multivariable-adjusted Cox proportional hazards regression for mLOY and death caused by cardiovascular disease, adjusted for the confounding variables outlined in table S2. The dependent variable in this primary model was mortality from any cardiovascular disease during follow-up, as indicated by a registered primary or secondary cause of death from the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD10) codes I00 to I99. This survival analysis established that mLOY in leukocytes is associated with death from overall cardiovascular disease in men [hazard ratio (HR) = 1.0054 per 1% increase in LOY, 95% confidence interval (CI) = 1.0022 to 1.0087, adjusted *P* = 0.0010, events = 6222). Using a penalized splines approach, the association could be illustrated over the full spectrum of LOY mosaicism (Fig. 2A), and a dose-response relationship was observed, indicating that mLOY might have a direct physiological effect. Exploratory Cox models using a binary LOY variable (defined from the spline analysis as described in fig. S6A) showed that men with LOY in >40% of leukocytes at study entry displayed 31% increased risk of dying from any disease of the circulatory system during follow-up (Fig. 2B, HR = 1.31, 95% CI = 1.02 to 1.70, adjusted *P* = 0.0382, events = 6222). To identify the specific cardiovascular outcomes that might be driving the overall association, we next investigated associations between mLOY and each of the ICD10 codes within I00 to I99 and with at least 200 observed events using analogous Cox models (table S3). The specific causes of death associated with mLOY included hypertensive heart disease (I11; HR = 3.48), heart failure (I50; HR = 1.76), congestive heart failure (I50.0; HR = 2.42), and aortic aneurysm and dissection (I71; HR = 2.76) (Fig. 2B). Because comorbidities are common among cardiovascular conditions, other registered cardiovascular-related causes of death showing association with mLOY (but normally not considered as primary causes of death) are summarized in fig. S6B. Next, cardiovascular diseases pooled on the basis of medical features and clinical resemblance were investigated using analogous survival models (table S4). In these extended analyses, there was no significant association between mLOY and death from diseases encompassing multiple etiologies of heart failure (HR = 1.65, 95% CI = 0.95 to 2.86, adjusted *P* = 0.0745, events = 1102). However, mLOY was associated with death caused by associated aortic and peripheral arterial diseases (HR = 2.79, 95% CI = 1.30 to 5.97, adjusted *P* = 0.0082, events = 279). Finally, corroborating prior analyses from the UK Biobank that used a shorter median (7-year) follow-up period (*8*), it was found that men with mLOY in >40% of leukocytes at study entry displayed a 41% increased risk of dying from any cause during the follow-up (Fig. 2B).

mLOY worsens the outcome of experimental heart failure in mice

To further investigate the relationship between mLOY and cardiac dysfunction, 12- to 16-week-old (4 weeks after BMT) male mice were subjected to transverse aortic constriction (TAC) surgery (Fig. 3A). Although TAC in mice does not completely replicate nonischemic heart failure, this widely used model produces cardiac dysfunction, left ventricle hypertrophy, and fibrosis phenotypes that are prevalent in the elderly (*13*). In agreement with observations in the older unchallenged mice, echocardiographic analysis revealed a greater progressive decline in cardiac function in mLOY mice compared with control mice after TAC (Fig. 3B and fig. S7A). The mLOY mice also displayed a marginally greater increase in the ratio of heart weight to tibia length, as well as an increase in the ratio of lung weight to tibia length that is indicative of lung congestion resulting from cardiac dysfunction (Fig. 3C). Transcripts encoding atrial natriuretic peptide A and the ratio of transcripts encoding myosin heavy chains β/α, a marker of advanced heart failure, were up-regulated in the hearts from mLOY mice compared with those from control mice (Fig. 3D). Histological analysis revealed greater interstitial and perivascular fibrosis in both the left ventricle and atrium of mLOY mice after TAC (Fig. 3E and fig. S7, B and C). Consistent with the increase in fibrosis, flow cytometric analysis of heart cells revealed that the number of MEF-SK4^{+} fibroblasts was higher in the hearts from TAC-treated mLOY mice compared with control (Fig. 3F). By contrast, there was no detectable difference in the number of cardiac endothelial cells or the average myocyte cross-sectional area between the experimental groups that had undergone TAC (Fig. 3, G and H), indicating that the predominant cellular effect of mLOY is on the fibroblast content of the heart.

Accelerated cardiac dysfunction could also be demonstrated in mice transplanted with lineage-negative bone marrow cells that were transduced with a gRNA that targets a different centromeric repeat in the Y chromosome (mLOY-gRNA2). The efficiency of

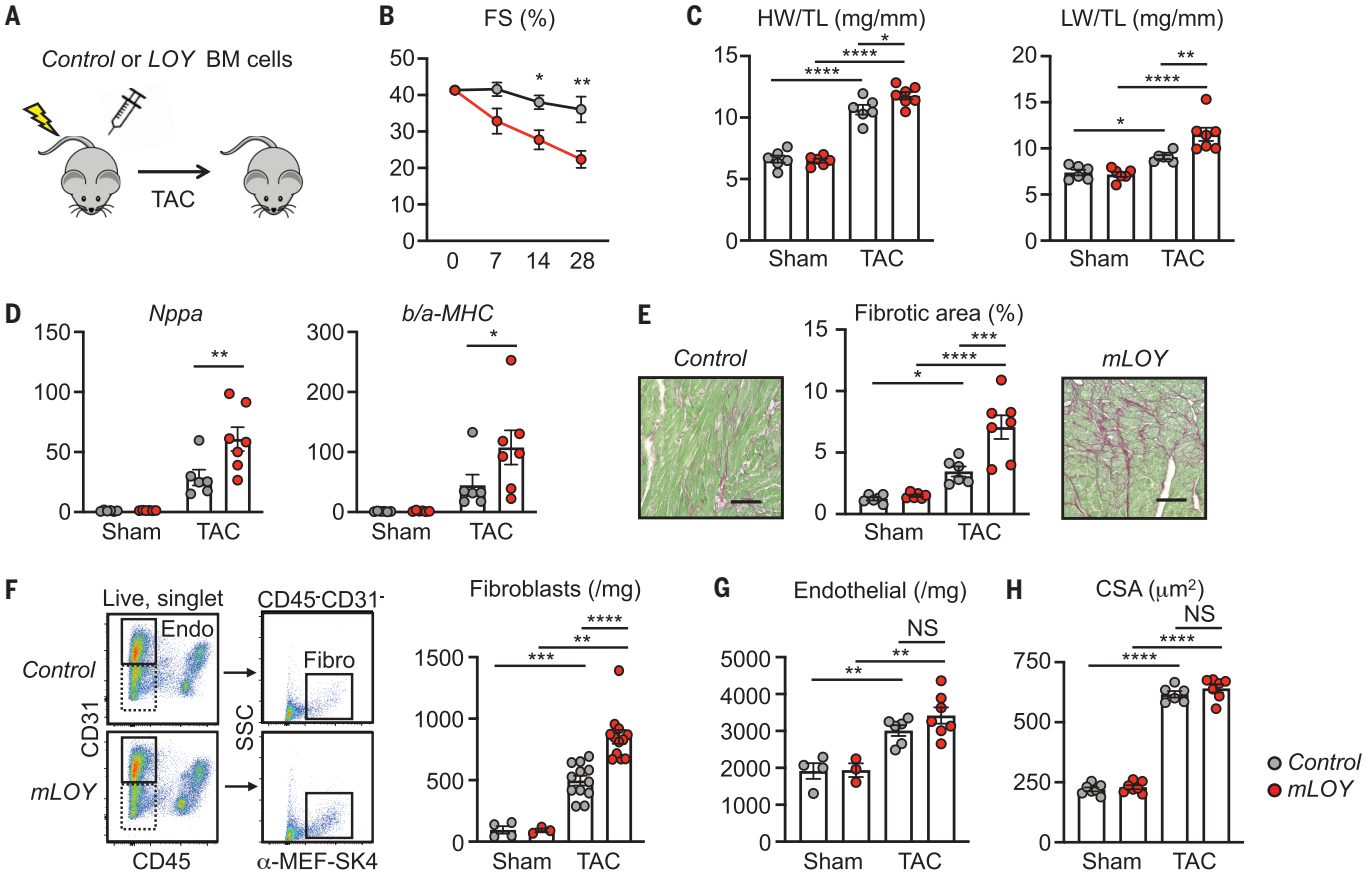


Fig. 3. Y chromosome deficiency in hematopoietic cells accelerates cardiac dysfunction in response to pressure overload. (**A**) Schematic of experimental study for assessing cardiac dysfunction of mLOY mice in the pressure overload model. mLOY mice were generated using LOY-gRNA1. At 4 weeks after BMT, mLOY mice or control mice were subjected to TAC. (**B**) Sequential echocardiographic analysis of mLOY and control mice after TAC at the indicated time points (control, $n = 6$; LOY, $n = 7$). (**C**) Heart weight (HW) and lung weight (LW) relative to tibial length (TL) in control sham mice and mice 4 weeks after TAC (control sham, $n = 6$; control mLOY, $n = 6$; control TAC, $n = 6$; mLOY TAC, $n = 7$). (**D**) Gene expression of the heart failure markers *Nppa* and *b/a-MHC* in heart tissue in sham mice and mice 4 weeks after TAC (control sham, $n = 6$; control mLOY, $n = 6$; control TAC, $n = 6$; mLOY TAC, $n = 7$). (**E**) Representative images and quantitative analysis (control sham, $n = 6$; control mLOY, $n = 6$; control TAC, $n = 6$; mLOY TAC, $n = 7$) of fibrotic area in heart sections in sham and 4-week after TAC. Scale bar, 100 μm. (**F** and **G**) Flow cytometric analysis of fibroblast (F) and endothelial cell (G) counts in heart tissue in sham mice and mice 4 weeks after TAC. Fibroblasts and endothelial cells are defined as $CD45^-CD31^-MEF\text{-}SK4^+$ and $CD45^-CD31^+$, respectively. The absolute numbers of cells were normalized by tissue weight ($n = 3$ to 4 per group in sham mice; $n = 6$ to 12 per group in TAC mice). (**H**) Mice with mLOY show comparable hypertrophic response of cardiac myocytes after TAC. Quantitative analysis of cross-sectional area of myocytes (CSA) in heart section at 4 weeks after TAC ($n = 6$ to 7 per group). Dots in all panels represent individual samples. Data are shown as mean ± SEM. Statistical analyses were performed using two-way repeated-measures ANOVA with Sidak's multiple-comparisons tests (B). Statistical analyses were performed using two-way ANOVA with Tukey's multiple-comparisons tests [(C), (D), (E), (F), and (G). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

Y chromosome ablation achieved by mLOY-gRNA2 was comparable but slightly less than that of mLOY-gRNA1 (Fig. 1B and fig. S8A). Transduction with mLOY-gRNA2 did not affect white blood cell counts, hemoglobin, or platelet counts (fig. S8B). However, TAC led to greater cardiac dysfunction, increased heart and lung weights, higher concentrations of serum brain natriuretic peptide (BNP), and higher numbers of cardiac fibroblasts in the mLOY-gRNA2–treated mice compared with control mice (fig. S8, C to F). These data corroborate the results with the mLOY-gRNA1 reagent and provide additional support for the finding that Y chromosome deletion in blood cells contributes to cardiac dysfunction and an accelerated fibrotic response.

Y chromosome deficiency modulates the transcriptional profile of cardiac macrophages

Myeloid cells typically display the greatest extents of Y chromosome deficiency in the blood of men (*12*). In the experimental model, analysis of the cardiac immune cell populations revealed higher numbers of $CCR2^+$ cardiac macrophages in TAC-treated hearts from mLOY-gRNA1 mice compared with those from control mice (fig. S9A), suggesting that cardiac macrophages derived from Y chromosome–deficient hematopoietic stem cells have altered functional properties. To test whether Y chromosome deficiency in myeloid cells accounts for the accelerated heart failure phenotype, anti–granulocyte receptor-1 (anti–GR-1) antibody, which blocks neutrophil and monocyte recruitment to injured tissue (*14*), was administered to the different experimental groups of mice. In mLOY mice that underwent TAC, treatment with anti–GR-1 antibody attenuated the accelerated cardiac dysfunction (fig. S9B) and reversed the elevations in heart weight and serum BNP (fig. S9, C and D). Treatment with anti–GR-1 antibody also reversed the increase of the number of cardiac fibroblasts observed in the mLOY mice, but it did not affect the quantity of vascular endothelial cells (fig. S9E). Collectively, these results suggest that loss of the Y chromosome in myeloid cells can largely account for the pathological cardiac phenotype in the mLOY mouse model.

To address the mechanistic aspects of mLOY-mediated cardiac dysfunction, single-cell RNA

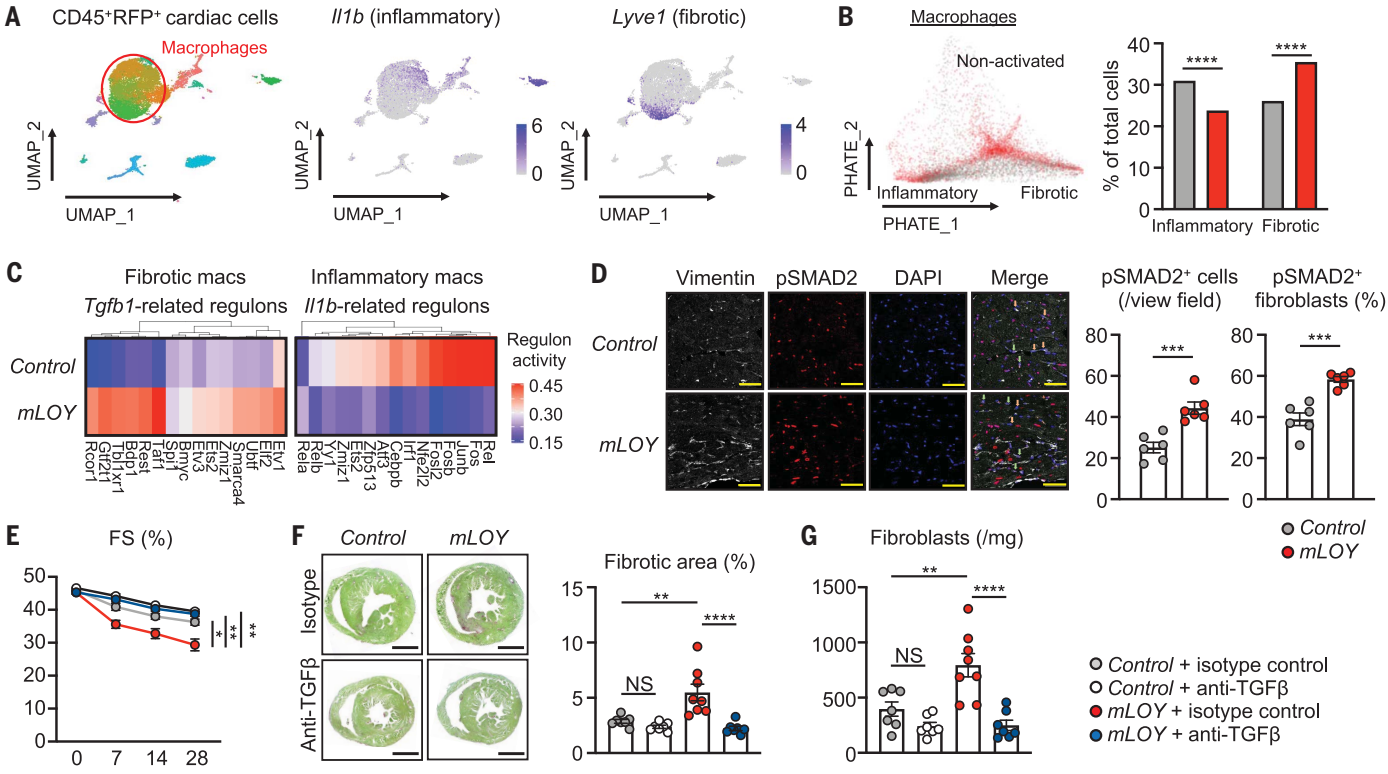


Fig. 4. Inhibition of TGFβ1 reverses cardiac dysfunction in mLOY mice after TAC. (**A**) Single-cell RNA sequencing from CD45$^+$RFP$^+$ cardiac cells 7 days after TAC shown by uniform manifold approximation and projection (UMAP) dimensionality reduction, with inflammatory and fibrotic macrophages and expression of *Il1b* and *Lyve1* highlighted. (**B**) PHATE dimensionality reduction showing cells separated from control (gray) or mLOY (red) samples, with nonactivated, inflammatory, and fibrotic phenotypes labeled. Quantification of the relative percentage of control or mLOY cells contained in the inflammatory or fibrotic macrophage clusters. (**C**) Heatmap of transcription factor regulons within inflammatory or fibrotic macrophages related to *Il1b* or *Tgfb1* expression, respectively, that was generated using SCENIC analysis of control and mLOY cells. (**D**) Activation of TGFβ1 signaling in the heart accessed by immunofluorescent staining of phosphorylated SMAD2 (pSMAD2). The number of pSMAD2$^+$ cells per view field and the percentage of pSMAD2$^+$ fibroblasts (right) are shown. Fibroblasts are defined as vimentin$^+$ cells (n = 6 fields per group). pSMAD2$^+$ and pSMAD2$^-$ fibroblasts are indicated by green and orange arrows, respectively. Scale bar, 50 μm. (**E**) At 4 weeks after BMT, mLOY mice or control mice were subjected to TAC. Anti-TGFβ1 antibody or isotype control was intraperitoneally injected every 3 days for 4 weeks. Shown is sequential echocardiographic analysis of mLOY and control mice after TAC operation at the indicated time points (n = 6 to 7 per group). (**F**) Representative images and quantitative analysis of fibrotic area in heart section at 4 weeks after TAC procedure (n = 6 to 8 per group). Scale bar, 1000 μm. (**G**) Flow cytometric analysis of fibroblast content in heart tissue at 4 weeks after TAC. The absolute numbers of cells were normalized by tissue weight (n = 7 to 8 per group). Dots in all panels represent individual samples. Data are shown as mean ± SEM. Statistical analyses were performed using chi-square test (B), unpaired Student's t test (D), two-way repeated-measures ANOVA with Sidak's multiple-comparisons test (E), and one-way ANOVA with Tukey's multiple-comparisons test [(F) and (G)]. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

sequencing was performed on BMT-derived (RFP$^+$) immune cells (CD45$^+$) from mLOY and control mice at 7 days after TAC. Seurat analysis of gene expression enabled the clustering of immune cell populations, and the reduced expression of Y chromosome–encoded genes could be detected in various clusters (fig. S10A). CD68$^+$ macrophages accounted for the largest portion of immune cells, which displayed polarization based on the expression of the marker genes *Il1b* and *Ccr2*^{hi}, which will henceforth define an inflammatory subpopulation, and the marker genes *Lyve1* and *Mrc1*, which will henceforth define a fibrotic subpopulation (Fig. 4A and fig. S10B). PHATE (potential of heat diffusion for affinity-based trajectory embedding) analysis to assess differentiation progression and branching revealed a trajectory from nonactivated macrophages to a continuum of inflammatory and fibrotic macrophages (fig. S10C), with the fibrotic macrophage subpopulation containing a greater percentage of mLOY cells and the inflammatory macrophage subpopulation containing a greater percentage of control cells (Fig. 4B). Consistent with the interpretation that mLOY promotes macrophage polarization toward a fibrotic phenotype, analysis with the SCENIC (single-cell regulatory network inference and clustering) regulatory gene network algorithm revealed that profibrotic regulons were enriched by the mLOY condition in the fibrotic macrophages (fig. S10D). Conversely, proinflammatory regulons were suppressed by the mLOY condition in the inflammatory macrophage subpopulation. Further analyses revealed that the mLOY condition promoted the enrichment of regulons specifically associated with transforming growth factor–β1 (TGFβ1) signaling in the fibrotic macrophage subpopulation, whereas regulons associated with interleukin-1β (IL-1β) signaling were down-regulated by mLOY in the inflammatory subpopulation (Fig. 4C). The mLOY condition promoted the expression of the transcript encoding TGFβ1 in the fibrotic macrophages, but not in the inflammatory macrophages, whereas the transcript encoding galectin-3, another secreted profibrotic protein, was up-regulated by the mLOY condition in both macrophage populations (fig. S10E).

The phenotypic transition of Y chromosome–deficient cardiac macrophages was also evident in a bulk transcriptome analysis of BMT-derived

(RFP$^+$) cardiac macrophages. As shown in the principal components analysis plots, mLOY macrophages isolated from heart displayed a transcriptomic profile distinct from that of control cardiac macrophages (fig. S11A). Gene set enrichment analysis (GSEA) revealed that differentially expressed transcripts related to "TGFβ signaling" were enriched in mLOY macrophages (fig. S11B). In addition, GSEA revealed that differentially expressed transcripts related to "TGFβ binding," including LTBP1, LTBP3, LTBP4, and thrombospondin, which facilitate the localization, secretion, and activation of TGFβ1, as well as other profibrotic factors, were up-regulated in macrophages in the mLOY condition (fig. S11C). In marked contrast, bulk or single-cell analysis revealed little or no differences in transcriptomes of cardiac monocytes, cardiac neutrophils, or blood neutrophils between the mLOY and control conditions (fig. S11, D to G).

Consistent with the transcriptome analyses, elevated TGFβ1 protein and SMAD2 phosphorylation in the myocardium could be detected by immunoblot analysis in the mLOY condition compared with control at 1 week after TAC (fig. S12, A and B). Elevated TGFβ1 protein in the mLOY condition could also be detected by fluorescence immunohistochemistry (fig. S12C). Although TGFβ1/macrophage colocalization was found to a greater extent in the mLOY condition, most of the TGFβ1 signal was localized to other cell types or matrix in the myocardium, consistent with its putative role in a feedforward signaling loop and its affinity for extracellular matrix. Furthermore, the total number of phosphorylated SMAD2$^+$ cells was higher in mLOY mice compared with control mice at 1 week after TAC, as confirmed by quantitative analysis of phosphorylated SMAD2$^+$ fibroblasts identified by co-immunostaining with vimentin (Fig. 4D).

TGFβ1 neutralization reverses the cardiac dysfunction observed in mLOY mice

To assess the role of profibrotic signaling in accelerated cardiac dysfunction caused by the experimental mLOY condition, mice that had undergone TAC were treated with control immunoglobulin G or a monoclonal antibody directed toward TGFβ1. Sequential echocardiographic analysis revealed that treatment with anti-TGFβ1 monoclonal antibody partly reversed the accelerated cardiac dysfunction observed in mLOY mice (Fig. 4E and fig. S12D). The anti-TGFβ1 antibody also reversed the mLOY-mediated increases in serum BNP concentration, heart weight, and lung congestion (fig. S12, E and F). Consistent with the recognized functions of TGFβ1 in promoting fibroblast proliferation and their conversion to myofibroblasts that produce higher amounts of matricellular proteins, the neutralizing anti-TGFβ1 antibody suppressed the increase in fibroblast number and extracellular matrix deposition that was associated with the mLOY condition but did not alter endothelial cell number (Fig. 4, F and G, and fig. S12G).

Discussion

Tissue fibrosis is a hallmark of aging and is estimated to contribute to 45% of deaths in industrialized countries (*13*). Myocardial fibrosis results from the activation of cardiac-resident fibroblasts and is often associated with heart failure, a major cause of mortality and morbidity in the elderly. Myocardial fibrosis can be triggered by bone marrow–derived macrophages that acutely infiltrate the heart in response to various forms of cardiac injury or progressively replace the cardiac-resident, yolk sac–derived macrophages with age (*15*, *16*). Here, we provide evidence in mouse models that supports a causal link between hematopoietic mLOY and age-dependent cardiac dysfunction and heart failure in men. We report that Y chromosome–deficient cardiac macrophages overactivate a profibrotic signaling network, leading to cardiac fibroblast proliferation and activation, excessive matrix production, and diminished heart function. The observation of profibrotic signaling in mLOY is unexpected compared with findings of how CHIP mechanistically contributes to cardiovascular disease. Whereas there is a high co-occurrence of mLOY with CHIP (*5*, *6*), and both are associated with cardiovascular disease mortality, including heart failure (*17*, *18*), CHIP appears to largely promote pathological processes through the overactivation of IL-1β/IL-6 inflammatory signaling in myeloid cells (*19*–*22*). It is increasingly recognized that chronic diseases are caused by a spectrum of inflammation- and fibrosis-driven events (*23*). However, the interrelationship between inflammation and fibrosis is not simply reciprocal in that chronic inflammation will promote fibrosis, whereas fibrosis can function in the resolution of inflammatory processes. Thus, the somatic mosaicism that develops in the hematopoietic system with age may give rise to a complex interplay of pro- and anti-inflammatory processes that can differentially affect disease development. Finally, our experimental studies also found that a neutralizing TGFβ1 antibody could reverse the pathological cardiac phenotypes caused by mLOY. In view of recent efforts to treat heart failure, idiopathic pulmonary fibrosis, and some cancers with antifibrotic approaches (*24*–*27*), men with mLOY could represent a patient subpopulation that exhibits a superior response to this class of therapeutic agents.

REFERENCES AND NOTES

1. M. A. Jobling, C. Tyler-Smith, *Nat. Rev. Genet.* **18**, 485–497 (2017).
2. L. A. Forsberg *et al.*, *Nat. Genet.* **46**, 624–628 (2014).
3. J. P. Dumanski *et al.*, *Science* **347**, 81–83 (2015).
4. D. J. Thompson *et al.*, *Nature* **575**, 652–657 (2019).
5. V. Ljungström *et al.*, *Leukemia* **36**, 889–891 (2022).
6. F. Zink *et al.*, *Blood* **130**, 742–752 (2017).
7. L. A. Forsberg *et al.*, *Nat. Genet.* **51**, 4–7 (2019).
8. E. Loftfield *et al.*, *Sci. Rep.* **8**, 12316 (2018).
9. J. P. Dumanski *et al.*, *Am. J. Hum. Genet.* **98**, 1208–1219 (2016).
10. S. Haitjema *et al.*, *Circ. Cardiovasc. Genet.* **10**, e001544 (2017).
11. S. Sano *et al.*, *J. Vis. Exp.* (152): (2019).
12. J. P. Dumanski *et al.*, *Cell. Mol. Life Sci.* **78**, 4019–4033 (2021).
13. D. C. Rockey, P. D. Bell, J. A. Hill, *N. Engl. J. Med.* **372**, 1138–1149 (2015).
14. J. M. Daley, A. A. Thomay, M. D. Connolly, J. S. Reichner, J. E. Albina, *J. Leukoc. Biol.* **83**, 64–70 (2008).
15. A. J. Rhee, K. J. Lavine, *Annu. Rev. Physiol.* **82**, 1–20 (2020).
16. Y. Wang *et al.*, *JCI Insight* **5**, e135204 (2020).
17. L. Dorsheimer *et al.*, *JAMA Cardiol.* **4**, 25–33 (2019).
18. B. Yu *et al.*, *J. Am. Coll. Cardiol.* **78**, 42–52 (2021).
19. J. J. Fuster *et al.*, *Science* **355**, 842–847 (2017).
20. S. Sano *et al.*, *Circ. Res.* **123**, 335–341 (2018).
21. S. Sano *et al.*, *J. Am. Coll. Cardiol.* **71**, 875–886 (2018).
22. Y. Yura *et al.*, *Circ. Res.* **129**, 684–698 (2021).
23. M. Wijsenbeek, V. Cottin, *N. Engl. J. Med.* **383**, 958–968 (2020).
24. T. E. Kang Jr. *et al.*, *N. Engl. J. Med.* **370**, 2083–2092 (2014).
25. G. A. Lewis *et al.*, *Nat. Med.* **27**, 1477–1482 (2021).
26. A. Fujiwara *et al.*, *Sci. Rep.* **10**, 10900 (2020).
27. H. Aghajanian *et al.*, *Nature* **573**, 430–433 (2019).

ACKNOWLEDGMENTS

We thank the participants and enablers of the UK Biobank study. This research was conducted using the UK Biobank Resource under application no. 61272 and enabled by the University of Virginia Flow Cytometry Core (RRID: SCR_017829) and the Genome Analysis and Technology Core (RRID:SCR_018883). LV.T11.Hosb8.Puro was provided by R. Förster at Hannover Medical School. **Funding:** This work was supported by the National Institutes of Health (NIH grants AG073249, AG072095, HL141256, HL139819, and HL142650 to K.W.; NIH grant HL152174 to S.S. and K.W.; Grant-in-Aid for Research Activity Start-up 21K20879 to S.S.; Grant-in-Aid for Scientific Research C 22K08162 to S.S.; and NIH Grant HL146056 to K.K.H.); the MSD Life Science Foundation (S.S.); the Ichiro Kanehara Foundation (S.S.); the Kenzo Suzuki Memorial Foundation (S.S); YOKOYAMA Foundation for Clinical Pharmacology (S.S.); the Cardiovascular Research Fund (S.S.); The Japanese Heart Failure Society (S.S.); the Osaka Medical Research Foundation for Intractable Diseases (S.S.); The European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grants 679744 and 101001789 to L.A.F.); the Swedish Research Council (grant 2017-03762 to L.A.F.); The Swedish Cancer Society (grant 20-1004 to L.A.F.); Kjell och Märta Beijers Stiftelse (L.A.F.); Marcus Borgströms stiftelse (J.H.); and the American Heart Association (grant 20POST35210098 to M.A.E.). **Author contributions:** S.S., K.K.H., L.A.F., and K.W. designed the experiments. S.S., K.H., H.O., Y.W., M.S., A.H., E.M.-Y., M.A.E., H.N.D.H., Y.Y., A.K., N.W.C., Yo.A., M.C.T., Yu.A., and J.C.M. performed the experiments. J.H., J.M., M.D., A.Z., T.F., J.S., and L.A.F. analyzed the UK Biobank data. S.S., H.O., L.A.F., and K.W. wrote the paper. All authors contributed to editing the final version of the paper. **Competing interests:** L.A.F. is cofounder and shareholder of Cray Innovation AB. The remaining authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Data were deposited into the National Center for Biotechnology Information's Gene Expression Omnibus database under accession numbers GSE205091 and GSE204732. **License information:**

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn3100
Materials and Methods
Figs. S1 to S12
Tables S1 to S7
References (*28*–*30*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 17 November 2021; accepted 6 June 2022
10.1126/science.abn3100

REPORT

SURFACE CHEMISTRY

Selectivity in single-molecule reactions by tip-induced redox chemistry

Florian Albrecht[1]*†, Shadi Fatayer[1,2]†, Iago Pozo[3]‡, Ivano Tavernelli[1], Jascha Repp[4], Diego Peña[3]*, Leo Gross[1]*

Controlling selectivity of reactions is an ongoing quest in chemistry. In this work, we demonstrate reversible and selective bond formation and dissociation promoted by tip-induced reduction-oxidation reactions on a surface. Molecular rearrangements leading to different constitutional isomers are selected by the polarity and magnitude of applied voltage pulses from the tip of a combined scanning tunneling and atomic force microscope. Characterization of voltage dependence of the reactions and determination of reaction rates demonstrate selectivity in constitutional isomerization reactions and provide insight into the underlying mechanisms. With support of density functional theory calculations, we find that the energy landscape of the isomers in different charge states is important to rationalize the selectivity. Tip-induced selective single-molecule reactions increase our understanding of redox chemistry and could lead to novel molecular machines.

Controlling selectivity in reactions is a central goal in chemistry. In solution, such control can be achieved by steering the valence electrons through the pH value or the electrochemical potential, for example. By these means, however, the reaction conditions are altered to such a degree that the basic mechanisms governing selectivity often remain elusive. The exploration of how external electrostatic fields and charge-state manipulation affect chemical bonding is still in its infancy (*1*).

Investigating surface chemistry by means of scanning tunneling microscopy (STM) and atomic force microscopy (AFM) offers the possibility to study basic chemical mechanisms under atomically well-defined conditions. The reaction itself can even be directly triggered with the tip of a scanning tunneling microscope at will (*2*, *3*). Spurred by advances in molecular characterization by STM (*4*) and AFM (*5*), novel tip-induced reactions and reaction mechanisms have been discovered (*6*–*11*). Typically, in tip-induced chemistry, precursors are designed on which specific bonds can be broken, dissociating designated masking groups (*3*). The demasking can in turn enable or cause other reactions, such as intermolecular bond formation (*3*, *11*, *12*), intramolecular bond formation (*7*), or skeletal rearrangements (*13*).

Control over configurational switching (*14*), bond formation and dissociation (*15*), and hydrogen tautomerization reactions (*16*–*18*) could be achieved by means of charge attachment and charge-state manipulation. By means of the electric field, control of configurational isomers (*19*) and control of the yield of a Diels-Alder reaction (*20*) were demonstrated. Furthermore, selectivity between molecular translation and desorption, controlled by inelastically tunneling electrons (*21*), and selective bond dissociation resulting from adsorbate-substrate bond alignment (*22*) were achieved. Even tip-controlled artificial molecular machines (*23*) have been demonstrated: for example, a molecule translated by molecular motors driven by inducing alternating conformational and configurational isomerization reactions (*24*).

Here, we showcase the potential of tip-induced electrochemistry for obtaining chemical selectivity in single-molecule reactions, that is, we demonstrate that multiple constitutional isomerization reactions can be controlled and selected by voltage pulses from the tip. Through selection of a specific voltage pulse, we formed different transannular covalent bonds.

As our molecular precursor, we synthesized 5,6,11,12-tetrachlorotetracene [**1** ($C_{18}H_8Cl_4$); see Fig. 1 and supplementary materials (SM), schemes S1 and S2, for details on the synthesis]. We deposited **1** by thermal sublimation on a Cu(111) substrate partly covered with NaCl islands of one- to three-monolayer (ML) thickness, at a sample temperature of $T \approx 10$ K. Experiments were performed at $T = 5$ K, on molecules on 2-ML NaCl on Cu(111), unless noted otherwise. All images reported were obtained with CO-functionalized tips (*5*). All AFM images are recorded in constant-height mode at a sample voltage, V, of 0 V.

Figure 2A shows an AFM image of **1**. The differences in brightness of the Cl atoms result from steric hindrance between neighboring Cl atoms, causing different adsorption heights (fig. S1). Using voltage pulses from the probe tip located above the molecule, we dissociated Cl atoms. We observed a threshold voltage of about +3.5 V, with tunneling currents on the order of $I = 1$ pA, for the dissociation of the first two Cl atoms from **1**. Often, the molecule also moved on the surface by a few nanometers when a voltage pulse was applied.

Figure 2B and fig. S2 show the partly dechlorinated intermediate **2** ($C_{18}H_8Cl_2$), with two Cl atoms of **1** dissociated (other partly dechlorinated intermediates that were observed are given in fig. S3 and table S1). The AFM image of **2** reveals a 10-membered ring on the dechlorinated side, exhibiting characteristic sharp and bright features above the triple bonds (*7*). This suggests that a retro-Bergman cyclization reaction (*7*) had taken place with intermediate **1a** as the transition structure. With voltage pulses of +4 to +4.5 V, we dissociated the remaining chlorine atoms in **2**, creating different structures with the chemical formula $C_{18}H_8$. We observed constitutional isomers ***3***, ***4***, and ***5*** (Fig. 2, C to E, respectively; see also figs. S4 to S6) and, in rare occasions, other isomers (fig. S7).

These molecules are highly strained and presumably very reactive, and none of them had been previously reported. Because of the inert NaCl surface and low temperature, they are stable under the conditions of our experiment for $|V| < 0.7$ V. The isomers differ in their central part, where they either exhibit a 10-membered ring (as in ***3***), a four-membered ring fused to an eight-membered ring (as in ***4***), or two fused six-membered rings, yielding a chrysene-like carbon backbone (as in ***5***). In most cases (62%), we found structure ***3*** directly after the dissociation of all chlorine atoms (table S2).

Figure 1 shows a plausible synthetic route for the formation of these strained isomeric hydrocarbons. At first, the cleavage of two Cl atoms from intermediate **2** probably generates sigma diradical **2a**, which could evolve through a Cope-type rearrangement to yield structure ***3***. This reaction then possibly proceeds in a transiently charged state (see the "Details on dehalogenations" section of the supplementary text). This would be in accordance with the mechanism previously proposed for enediyne cyclizations, which is facilitated by the formation of radical-anionic species (*25*). While ***3*** is a plausible resonance structure with two cumulene moieties, alternative

[1]IBM Research Europe – Zurich, 8803 Rüschlikon, Switzerland. [2]Applied Physics Program, Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia. [3]Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CiQUS) and Departamento de Química Orgánica, Universidade de Santiago de Compostela, 15782-Santiago de Compostela, Spain. [4]Institute of Experimental and Applied Physics, University of Regensburg, 93053 Regensburg, Germany.
***Corresponding author. Email: fal@zurich.ibm.com (F.A.); diego.pena@usc.es (D.P.); lgr@zurich.ibm.com (L.G.)**
†These authors contributed equally to this work.
‡Present address: Chemistry Research Laboratory (CRL), Department of Chemistry, University of Oxford, Oxford OX1 3TA, UK.

structures combining enyne and cumulene groups (***3′***) or two enynes (***3″***) within the central 10-membered ring can also be considered. The transannular C–C bond formation between the radicals in structure ***3″*** would lead to the formation of diyne **4**, whereas the C–C bond formation between radicals in ***3′*** would afford the chrysene-based bisaryne **5**.

First, we characterize on the NaCl surface the products **3**, **4**, and **5**. The STM-measured maps of electronic resonances are shown in Fig. 3 and fig. S8, accompanied by density functional theory (DFT) calculations. We find that **4** and **5** are charge neutral, with the shape of the imaged frontier orbital densities, of the highest occupied and lowest unoccupied molecular orbital (HOMO and LUMO, respectively), in agreement with the DFT calculations (Fig. 3, A to K, and fig. S9). For **4**, we did not observe the positive ion resonance (HOMO) for voltages up to −2 V (fig. S10). For **3**, the experiment shows that the molecule is in its anionic charge state. This is consistently indicated by the scattering of interface-state electrons (fig. S11), Kelvin probe force spectroscopy (table S3), and STM maps of the electronic resonances (Fig. 3, B and C) and their comparison with theory (figs. S12 and S13). The identification of structure ***3*** as a radical anion agrees with the expected tendency of sigma radicals to be reduced, in contrast to closed-shell compounds **4** and **5**.

Moreover, our measurements reveal that the anion of ***3*** undergoes a Jahn-Teller distortion that is also found by B3LYP-based DFT (Fig. 3P) and explains the symmetry breaking with respect to the long molecular axis that is observed for both electronic resonances (Fig. 3, B and C). Structurally, the Jahn-Teller distortion of the anion $\boldsymbol{3}^{-1}$ features an inward bend on one long side of the 10-membered ring (Fig. 3P) and can be observed in high-resolution AFM images (Fig. 3M and fig. S4). The first electronic resonance at negative bias shows two lobes of increased orbital density on the side of the inward bend (Fig. 3, B and H, and fig. S12). Whereas the first electronic resonance at positive bias exhibits increased density on the side opposite to the inward bend (Fig. 3, C and I, and figs. S12 and S13), in excellent agreement of experiment and theory.

Next, we study tip-induced reactions between **3**, **4**, and **5**. When we applied pulses of relatively large bias, $V > +2.5$ V with I on the order of 10 pA, we could transform molecules between all of these three structures—with limited control of the outcome, however. A rearrangement after a bias pulse of $V = +2.5$ V mostly resulted in structure **5** (~50% of the attempts) and less often in structures **4** or **3** (25% each). Structure **5** was stable for $|V| < 2$ V. However, voltage pulses of $|V| < 2$ V, when applied to **3** and **4**, resulted in different reactions, depending on the applied voltage.

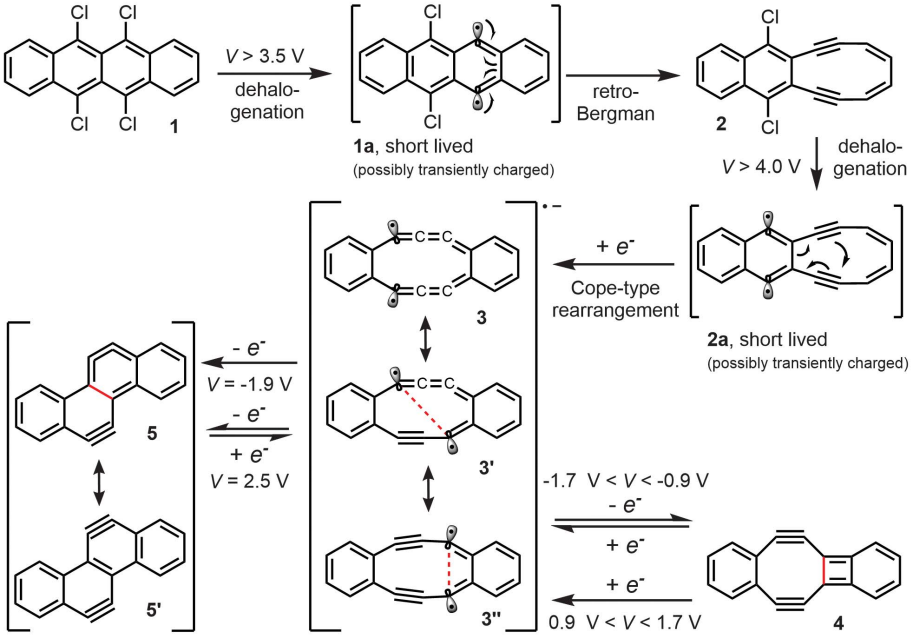


Fig. 1. Reaction scheme. Tip-induced reactions promoted by voltage pulses are indicated with the respective sample voltages V. For **3** and **5**, different resonance structures are displayed. The intermediates **1a** and **2a** might be transiently charged. We also observed partly dehalogenated intermediates other than **2** (see fig. S3), indicating several pathways for the initial formation of **3**, **4**, and **5**.

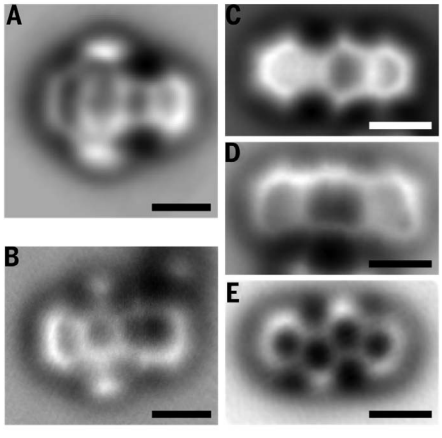


Fig. 2. AFM images of precursor, intermediate, and products. Constant-height, CO-tip AFM images of precursor **1** (**A**), intermediate **2** (**B**), and products **4** (**C**), **3** (**D**), and **5** (**E**) on NaCl. (A), (B), and (D) on 2-ML NaCl, and (C) and (E) on 1-ML NaCl. Scale bars, 0.5 nm. For images at different tip-height offsets and imaging parameters, see figs. S1, S2, and S4 to S6.

Histograms showing the outcome of voltage pulses with currents of 0.5 pA applied above ***3*** and **4** as initial structures are shown in Fig. 4, A and B, respectively (see also figs. S14 and S15). Our findings are as follows: (i) At $|V| < 0.7$ V, ***3*** and **4** are stable. (ii) At -1.7 V $< V < -0.9$ V, bidirectional switching occurs between ***3*** and **4**. That is, ***3*** can be converted into **4** (transition labeled α) and vice versa (transition β), and only with a small probability (<10%) structure **5** is formed. (iii) At $+0.9$ V $< V < +1.7$ V, unidirectional switching from **4** to ***3*** occurs. That is, **4** is converted into ***3*** (transition γ), but ***3*** is stable at these voltages. (iv) At $V \approx -1.9$ V, structure ***3*** is transformed into **5** with a high yield, and structure **4** is transformed into either ***3*** or **5**.

Switching between the structures was possible on 1-, 2-, and 3-ML NaCl islands with similar threshold values. The data shown in Fig. 4 were measured on 2-ML NaCl. The observed voltage dependence allows us to select outcomes of tip-induced rearrangements. We can select which transannular C–C bond is formed within the 10-membered ring of structure ***3***. We dominantly generated structure **4** with pulses in the range of −1.1 to −1.7 V, and structure **5** with $V = -1.9$ V, demonstrating selectivity in single-molecule constitutional isomerization reactions.

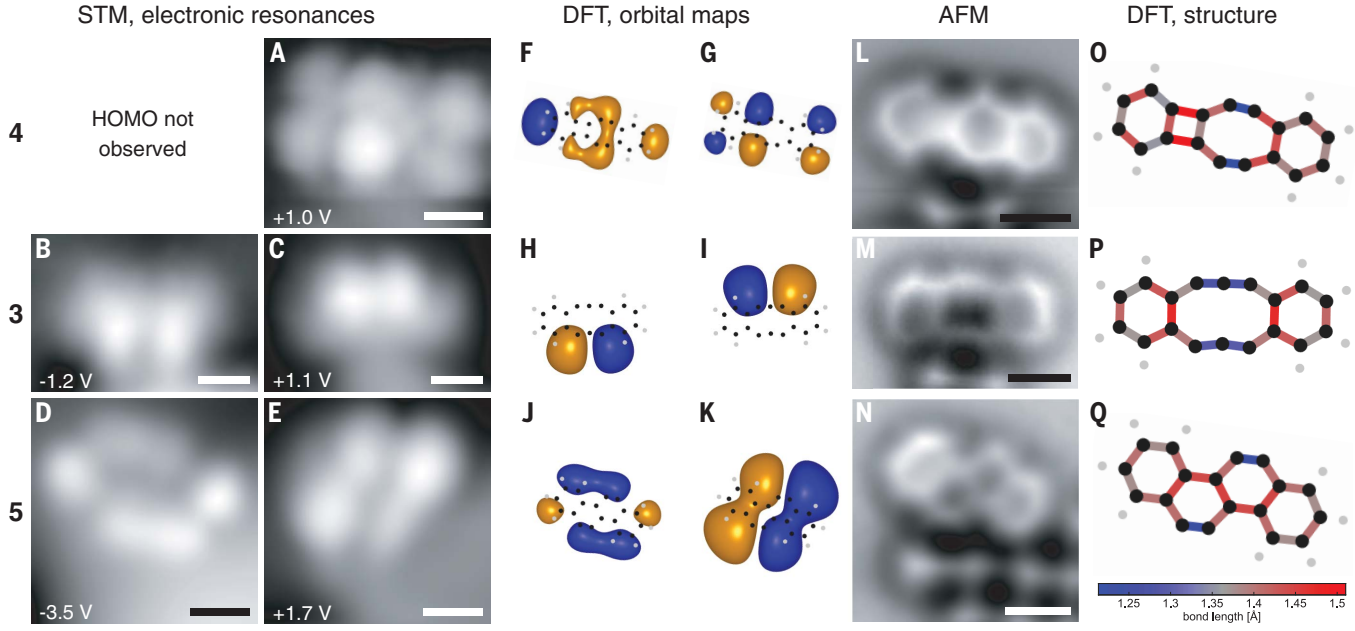
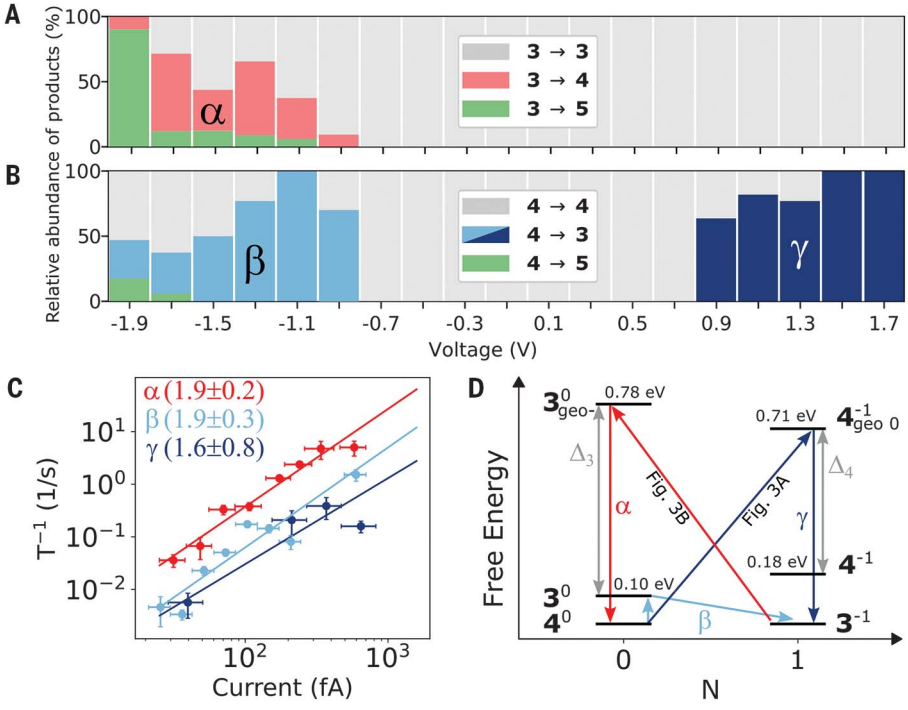


Fig. 3. Characterization of products by STM, DFT, and AFM. (**A** to **E**) STM images at the first electronic resonances at negative and positive sample voltage. (**F** to **K**) Simulated orbital maps of the frontier orbitals, considering the finite resolution of the tip (see fig. S9). (**L** to **N**) AFM images. All experimental images were recorded on 2-ML NaCl, with the molecules next to 3-ML NaCl islands. (**O** to **Q**) DFT-calculated structures on NaCl with the bond length indicated by color. The DFT calculations of **3** are shown in the anionic charge state, those for **4** and **5** are shown in the neutral charge state. Scale bars, 0.5 nm.

Fig. 4. Tip-induced transitions between 3 and 4. Histograms (**A** and **B**) show the outcome in dependence of V applied for 10 s at constant current of 0.5 pA above molecule **3** (A) and **4** (B) on 2-ML NaCl. (**C**) Reaction rate as a function of tunneling current. Transitions α and β were probed at V = −1.3 V, transition γ was probed at V = 1.1 V. The extracted slopes are 1.86 ± 0.18 for α, 1.90 ± 0.26 for β, and 1.61 ± 0.76 for γ. (**D**) Energy level diagram for $V \approx 0$ V, associated with the observed charging and structural transitions between **3** and **4**, assuming equal free energies of $\mathbf{4}^{0}$ and $\mathbf{3}^{-1}$. N denotes the number of excess electrons.



For the transitions between **3** and **4**, we investigated the respective reaction rates (Fig. 4C). These were measured at V = −1.3 V for transitions α (from **3** to **4** at negative V) and β (from **4** to **3** at negative V) and at V = +1.1 V for γ (from **4** to **3** at positive V), as a function of current by using different tip heights. The slopes of the linear fits in the double logarithmic plot are 1.86 ± 0.18 for α, 1.90 ± 0.26 for β, and 1.61 ± 0.76 for γ. This indicates that transitions α and β are two-electron processes. For γ, the error is too large to differentiate between a one-electron and a two-electron process.

Figure 4D visualizes the transitions between **3** and **4**. The transitions α and γ coincide with the onset of ionic resonances of the initial structures, as probed by STM (Fig. 3, B and A, respectively), suggesting that these transitions involve (de)charging the initial structure. Note that at electronic resonance, the charge state

is transiently changed by charge transfer between tip and molecule. After a typical lifetime on the order of a few picoseconds on 2-ML NaCl, the molecular charge ground state is restored by charge transfer between molecule and metal substrate (*4*). The structural relaxation that follows a charge transition can oscillate for several tens of picoseconds (*26*).

For Fig. 4D, we calculated on a NaCl surface the ground state energies of **3** and **4** in different charge states and the related relaxation energies Δ (see SM) (*27*). The relatively large energy of the intermediate $\mathbf{3}^{\mathbf{0}}_{\mathbf{geo-}}$ for the transition from $\mathbf{3}^{-\mathbf{1}}$ is rationalized by the Jahn-Teller distortion of $\mathbf{3}^{-\mathbf{1}}$. The calculated energies for the charge transitions are in good agreement with the resonances measured by STM when assuming similar energies for $\mathbf{4}^{\mathbf{0}}$ and $\mathbf{3}^{-\mathbf{1}}$. A partial voltage drop across NaCl of ~20% can be considered for this junction geometry (*27*).

The observed two-electron process for α indicates that in addition to the charge transition, a second charge carrier is needed to provide additional energy in an inelastic electron tunneling (IET) process (*28*). In contrast, transition β involves no charging of the initial structure $\mathbf{4}^{0}$, but probably only by IET processes a transition to $\mathbf{3}^{\mathbf{0}}$ occurs, which is subsequently charged from the substrate to $\mathbf{3}^{-\mathbf{1}}$, the charge ground state of structure **3**. In transition γ, the LUMO of **4**, shown in Fig. 3G, is transiently occupied. This orbital exhibits a nodal plane along the long axis of the molecule and therefore is an antibonding state with respect to the central bond of **4**, facilitating its rupture (*29*) and thus the transition to **3**. Note that at small currents of $I \approx 0.5$ pA, the reaction rates are on the order of minutes, and thus orbital density images can be obtained (Fig. 3, A to C). Our results indicate that these reactions are triggered by electron attachment rather than by the electric field alone. That the effect of the latter is small can be rationalized by the reaction coordinates, that is, the molecular plane and the movement of atoms, being parallel to the substrate, whereas the electric field applied by the tip is orthogonal with respect to the substrate and the reaction coordinates.

On-surface calculations (fig. S16) show that in both the neutral and negative charge state, **5** has a lower energy than **3** and **4**, explaining the observed dominant switching to **5** at larger biases and its stability for $|V| < 2.0$ V. Gas-phase calculations (fig. S16) indicate that the reaction barriers (both in neutral and negative charge states) to **5** are higher than the ones between **3** and **4**, explaining the increased voltages needed to switch to **5**, compared with switching between **3** and **4**. The high barrier between **4** and **5** in the neutral and negative charge states (fig. S16) suggests that the observed transformation from **4** to **5** at $V = -1.9$ V (Fig. 4B) proceeds via structure **3** as intermediate.

The switching at higher bias $|V| > 2.0$ V and increased currents on the order of 10 pA is less controlled and more challenging to understand, and its detailed description is beyond the scope of this work. Because of the large bias, several successive and branching charge transitions and structural rearrangements must be considered. Also, current-induced catalytic reduction of the barrier might play a role (*30*). In addition, the switching between all three structures at $V = 2.5$ V could include higher excited states, for example, occupation of the LUMO+1, which for **5** is accessed at $V = 2.5$ V (fig. S17). It might proceed via dianionic charge states, which show similar ground state energies for all three structures (fig. S16), and the transient occupation of the dianion is observed for **3** by resonant tunneling in Fig. 3C at $V = 1.1$ V.

Our experiment shows that for a molecule on a surface, several chemical transformations between multiple constitutional isomers can be controlled by tip-induced redox chemistry. With different voltages and polarities, we selectively activated one, two, or all three transitions between three different isomers (**3**, **4**, and **5**). We demonstrated directed and reversible switching between two non–ground state isomers (**3** and **4**) and selectively formed transannular covalent bonds deliberately transforming **3** to either **4** or **5**. We learned that the selectivity of the reactions is facilitated by changes of the energy landscape as a function of the transient charge state, accessed by the bias applied. The charge ground states can, in general, be adjusted via the substrate's work function (*31*).

The insights obtained in redox reactions studied by tip-induced electrochemistry will be useful to better understand redox reactions that are important in organic synthesis (*32*) and nature (*33*). For future artificial molecular machines (*23*), controlled, reversible, and selected switching between more than two different constitutional isomers, as demonstrated in our work, could enable novel functionalities. In addition, increased workload and operation at elevated temperatures could be facilitated by the relatively high energy barriers, on the order of 1 eV, involved in constitutional isomerization reactions.

REFERENCES AND NOTES

1. S. Ciampi, N. Darwish, H. M. Aitken, I. Díez-Pérez, M. L. Coote, *Chem. Soc. Rev.* **47**, 5146–5164 (2018).
2. B. Stipe *et al.*, *Phys. Rev. Lett.* **78**, 4410–4413 (1997).
3. S.-W. Hla, L. Bartels, G. Meyer, K.-H. Rieder, *Phys. Rev. Lett.* **85**, 2777–2780 (2000).
4. J. Repp, G. Meyer, S. M. Stojković, A. Gourdon, C. Joachim, *Phys. Rev. Lett.* **94**, 026803 (2005).
5. L. Gross, F. Mohn, N. Moll, P. Liljeroth, G. Meyer, *Science* **325**, 1110–1114 (2009).
6. J. Repp, G. Meyer, S. Paavilainen, F. E. Olsson, M. Persson, *Science* **312**, 1196–1199 (2006).
7. B. Schuler *et al.*, *Nat. Chem.* **8**, 220–224 (2016).
8. J. N. Ladenthin *et al.*, *Nat. Chem.* **8**, 935–940 (2016).
9. E. Kazuma, J. Jung, H. Ueba, M. Trenary, Y. Kim, *Science* **360**, 521–526 (2018).
10. S. Kawai *et al.*, *Sci. Adv.* **6**, eaay8913 (2020).
11. Q. Zhong *et al.*, *Nat. Chem.* **13**, 1133–1139 (2021).
12. K. Kaiser *et al.*, *Science* **365**, 1299–1301 (2019).
13. N. Pavliček *et al.*, *Nat. Chem.* **10**, 853–858 (2018).
14. B.-Y. Choi *et al.*, *Phys. Rev. Lett.* **96**, 156106 (2006).
15. F. Mohn *et al.*, *Phys. Rev. Lett.* **105**, 266102 (2010).
16. P. Liljeroth, J. Repp, G. Meyer, *Science* **317**, 1203–1206 (2007).
17. W. Auwärter *et al.*, *Nat. Nanotechnol.* **7**, 41–46 (2011).
18. T. Kumagai *et al.*, *Nat. Chem.* **6**, 41–46 (2014).
19. M. Alemani *et al.*, *J. Am. Chem. Soc.* **128**, 14446–14447 (2006).
20. A. C. Aragonès *et al.*, *Nature* **531**, 88–91 (2016).
21. J. I. Pascual, N. Lorente, Z. Song, H. Conrad, H.-P. Rust, *Nature* **423**, 525–528 (2003).
22. K. Anggara *et al.*, *Nat. Commun.* **7**, 13690 (2016).
23. S. Erbas-Cakmak, D. A. Leigh, C. T. McTernan, A. L. Nussbaumer, *Chem. Rev.* **115**, 10081–10206 (2015).
24. T. Kudernac *et al.*, *Nature* **479**, 208–211 (2011).
25. I. V. Alabugin, M. Manoharan, *J. Am. Chem. Soc.* **125**, 4495–4509 (2003).
26. T. L. Cocker, D. Peller, P. Yu, J. Repp, R. Huber, *Nature* **539**, 263–267 (2016).
27. S. Fatayer *et al.*, *Nat. Nanotechnol.* **13**, 376–380 (2018).
28. P. A. Sloan, R. E. Palmer, *Nature* **434**, 367–371 (2005).
29. A. Erpenbeck, Y. Ke, U. Peskin, M. Thoss, *Phys. Rev. B* **102**, 195421 (2020).
30. A. A. Dzhioev, D. S. Kosov, F. von Oppen, *J. Chem. Phys.* **138**, 134103 (2013).
31. F. E. Olsson, M. Paavilainen, M. Persson, J. Repp, G. Meyer, *Phys. Rev. Lett.* **98**, 176803 (2007).
32. M. Yan, Y. Kawamata, P. S. Baran, *Chem. Rev.* **117**, 13230–13319 (2017).
33. H. Beinert, R. H. Holm, E. Münck, *Science* **277**, 653–659 (1997).

ACKNOWLEDGMENTS

We thank R. Allenspach, S. Mishra, E. Guitián, and Z. Zhu for discussions. For some of the calculations, this research used the resources of the Supercomputing Laboratory at King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia. **Funding:** This work was supported by the ERC Synergy Grant MolDAM (951519), ERC Consolidator Grant AMSEL (682144), European FET-OPEN project SPRING (863098), Spanish Agencia Estatal de Investigación (PID2019-107338RB-C62, PID2019-110037GB-I00, and PCI2019-111933-2), Xunta de Galicia (Centro de Investigación de Galicia accreditation 2019–2022, ED431G 2019/03), and the European Regional Development Fund (ERDF). **Author contributions:** On-surface experiments: F.A., S.F., and L.G. Synthesis of precursor molecules: I.P. and D.P. DFT calculations: S.F., I.T., I.P., F.A., and L.G. Writing: L.G., F.A., S.F., I.P., I.T., J.R., and D.P. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo6471
Supplementary Text
Figs. S1 to S17
Tables S1 to S3
References (*34–49*)

Submitted 23 February 2022; accepted 24 May 2022
10.1126/science.abo6471

SOLAR CELLS

Efficient and stable perovskite-silicon tandem solar cells through contact displacement by MgF_x

Jiang Liu[1]*, Michele De Bastiani[1], Erkan Aydin[1], George T. Harrison[1], Yajun Gao[1], Rakesh R. Pradhan[1], Mathan K. Eswaran[1], Mukunda Mandal[2], Wenbo Yan[1], Akmaral Seitkhan[1], Maxime Babics[1], Anand S. Subbiah[1], Esma Ugur[1], Fuzong Xu[1], Lujia Xu[1], Mingcong Wang[1], Atteq ur Rehman,[1] Arsalan Razzaq[1], Jingxuan Kang[1], Randi Azmi[1], Ahmed Ali Said[1], Furkan H. Isikgor[1], Thomas G. Allen[1], Denis Andrienko[2], Udo Schwingenschlögl[1], Frédéric Laquai[1], Stefaan De Wolf[1]*

The performance of perovskite solar cells with inverted polarity (p-i-n) is still limited by recombination at their electron extraction interface, which also lowers the power conversion efficiency (PCE) of p-i-n perovskite-silicon tandem solar cells. A MgF_x interlayer with thickness of ~1 nanometer at the perovskite/C_{60} interface favorably adjusts the surface energy of the perovskite layer through thermal evaporation, which facilitates efficient electron extraction and displaces C_{60} from the perovskite surface to mitigate nonradiative recombination. These effects enable a champion open-circuit voltage of 1.92 volts, an improved fill factor of 80.7%, and an independently certified stabilized PCE of 29.3% for a monolithic perovskite-silicon tandem solar cell ~1 square centimeter in area. The tandem retained ~95% of its initial performance after damp-heat testing (85°C at 85% relative humidity) for >1000 hours.

Integrating high-performance wide-bandgap perovskite solar cells onto silicon solar cells can lead to very high power conversion efficiencies (PCEs) by minimizing carrier thermalization losses (*1*–*6*). Although initial research explored n-i-p tandems, recent work has focused on the p-i-n configuration, in which the n-type electron-collecting contact faces sunward (*7*–*9*), and on improving performance through device optics and optimizing perovskite composition (*10*–*15*). More recently, attention has turned to the interface between the perovskite and the hole transport layer (HTL) to reduce voltage losses. Approaches include molecular passivation of NiO_x (*16*, *17*) and the use of self-assembled monolayers (SAMs) such as 2PACz and Me-4PACz, anchored on oxides, to reduce open-circuit voltage (V_{oc}) losses (*18*, *19*).

Despite this progress, state-of-the-art PSCs, especially those incorporating wider-bandgap perovskites (e.g., ~1.68 eV as frequently used for tandem applications), have an undesirably large V_{oc} deficit when compared to the theoretical radiative limit. This problem mainly stems from substantial charge carrier recombination and an energy level mismatch at the perovskite interface with the electron transport layer (ETL) (*20*–*22*), which most commonly consists of evaporated C_{60}. Inserting an ultrathin LiF layer at the perovskite/C_{60} interface alleviates this issue to a certain extent, yet this may result in reduced device stability, usually attributed to the deliquescent behavior and high ion diffusivity of Li salts (*19*, *23*–*25*). Two-dimensional (2D) perovskites and some fullerene derivatives prepared by solution processes have been previously used to passivate the perovskite/C_{60} rear interface in single-junction p-i-n PSCs (*26*, *27*). However, in the p-i-n tandem configuration, the perovskite/C_{60} interface faces sunward, which demands interfacial layers with a high transparency, high stability, and good thickness control. To this end, we systematically investigated alternative evaporated metal fluorides (such as NaF, CaF_x, and MgF_x) as the interlayer at the perovskite/C_{60} interface. We demonstrated that the charge transport and recombination interfaces could be carefully tuned with MgF_x interlayers, enabling a certified stabilized PCE of 29.3%.

We fabricated monolithic perovskite-silicon tandem solar cells from silicon heterojunction bottom cells using crystalline silicon (c-Si) wafers with double-side texture (Fig. 1A) to reduce the front reflection and improve light trapping in our devices (*8*, *16*). We verified the ultrathin nature of the fluoride-based interlayers, inserted at the electron-selective top contact, with cross-sectional high-resolution scanning transmission electron microscopy (HR-STEM; Fig. 1B and figs. S1 and S2). The magnified STEM images and energy-dispersive x-ray (EDX) spectroscopy mapping clearly outline the perovskite/MgF_x/C_{60}/SnO_2/IZO top contact structure, identifying the presence of a ~15-nm C_{60} layer and a ~20-nm SnO_2 layer. The latter acts as a buffer against damage from sputtering of the indium zinc oxide (IZO) transparent top electrode (*28*). We note that after perovskite deposition, all subsequently deposited films were obtained by vapor deposition techniques that yield highly accurate and reproducible layer thicknesses. For instance, because the fluoride-based film is thermally evaporated, the resulting interlayer is highly uniform in thickness and less affected by the surface roughness of the underlying perovskite, unlike typical solution-processed interlayers (*27*).

We investigated the energy-level alignment of our perovskite layers with LiF, NaF, CaF_x and MgF_x overlayers by ultraviolet photoemission spectroscopy (UPS) and low-energy inverse photoemission spectroscopy (LE-IPES) for occupied and unoccupied states, respectively. As shown in Fig. 1C, the work function (WF) of the bare perovskite is ~4.97 eV. By coating the perovskite with a thin fluoride-based layer, the WF systematically shifts toward smaller values. Both MgF_x and CaF_x caused a larger WF shift than did LiF and NaF (Fig. 1C and fig. S3). With the presence of metal fluoride interlayers, the valence band maximum of the perovskite, determined with a Gaussian fitting method (*29*), was lowered relative to its Fermi level (E_F); this implies that the metal fluorides caused a downward band bending at the perovskite interface that favored electron extraction.

Kelvin probe force microscopy measurements conducted in an ambient environment confirmed the trend of the UPS results (fig. S4) (*30*); that is, the MgF_x and CaF_x samples displayed a larger WF shift than did the LiF and NaF samples (fig. S5). To further evaluate such band bending as a function of ETL thickness, we conducted additional UPS/LE-IPES measurements (figs. S6 and S7) that allowed us to map out the band structure at the perovskite/ETL interface. The perovskite/C_{60} sample displayed negligible band bending (fig. S6), which is consistent with previous work on the $MAPbI_3$/C_{60} interface (*31*). However, the presence of a MgF_x interlayer led to energy band bending at the perovskite surface (Fig. 1D). Also, the lowest unoccupied molecular orbital (LUMO) of the C_{60} layer bent down toward the perovskite interface, which implies that the MgF_x layer promoted the formation of electron-selective contacts with low interfacial resistance (*32*).

Moreover, the MgF_x interlayer also displaced C_{60} from the perovskite surface, thus suppressing interface recombination (see below). The thinness of the metal fluoride interlayer (thickness 0.5 to 1.5 nm) ensures that collected electrons can reach the LUMO of the C_{60} layer through quantum mechanical tunneling or via pinholes, thus enabling the selective extraction of electrons. Once the electrons transferred to C_{60}, they became the majority charge carriers and were easily transported through the C_{60} layer and collected by the SnO_2/IZO transparent electrode. Furthermore, x-ray photoelectron spectroscopy (XPS) results (fig. S8) showed that the evaporated ultrathin (~1 nm) MgF_x films strongly deviated from their bulk stoichiometric (x = 2) composition,

[1]KAUST Solar Center, Physical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia. [2]Max Planck Institute for Polymer Research, 55128 Mainz, Germany.
*Corresponding author. Email: jiang.liu@kaust.edu.sa (J.L.); stefaan.dewolf@kaust.edu.sa (S.D.W.)

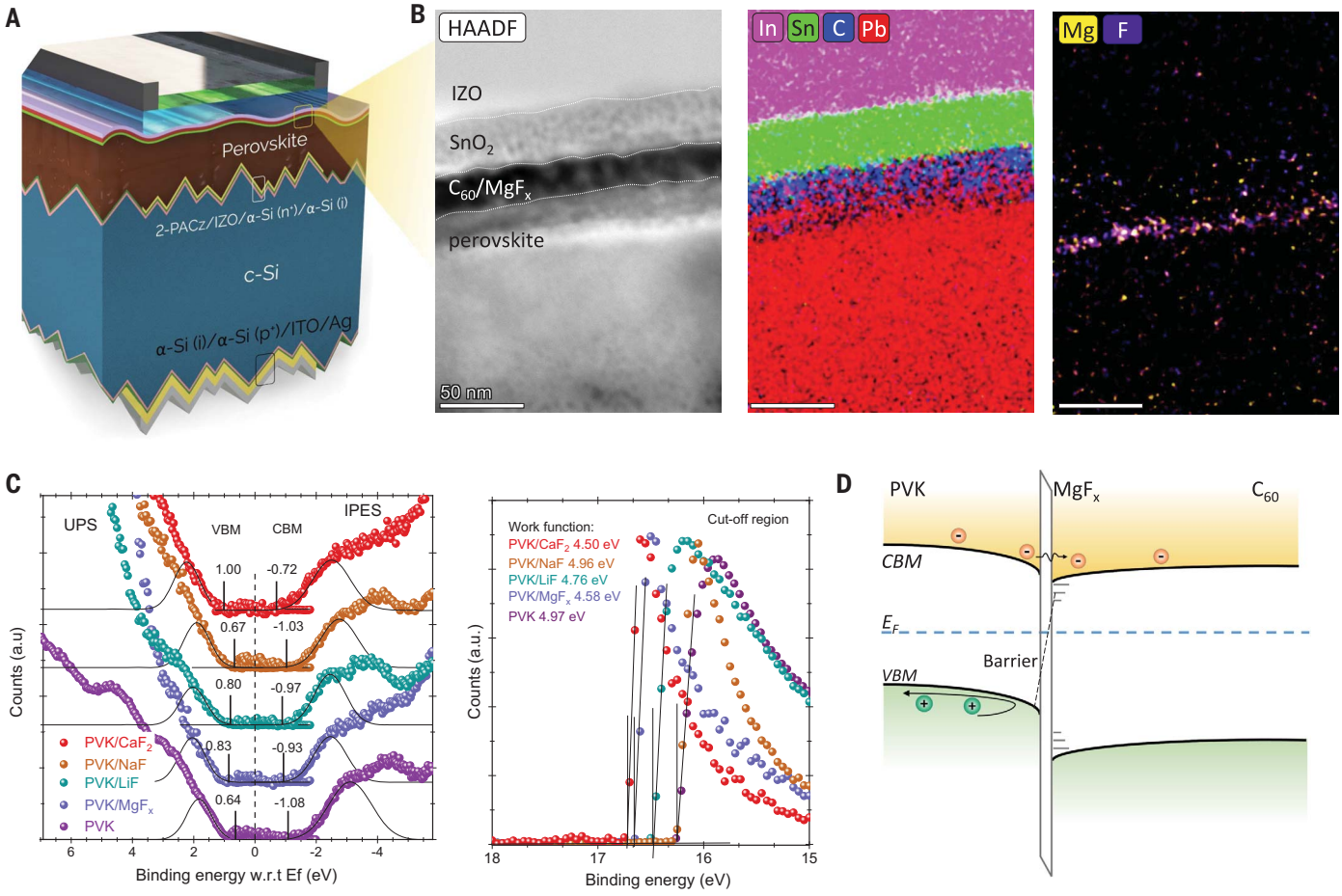


Fig. 1. Interface structure and electronic properties. (**A**) Schematic of the monolithic perovskite/silicon tandem solar cell built from a double-side textured silicon heterojunction cell. (**B**) High-angle annular dark-field (HAADF) image in HR-STEM and the corresponding EDX mapping at the ETL side. (**C**) Valence band and photoelectron cut-off region of the perovskite, and perovskite/1 nm-interlayer using UPS and IPES spectra. The conduction band minimum (CBM) and valence band maximum (VBM) were determined by fitting the spectral edge using Gaussian function. All samples were deposited on IZO/2PACz coated c-Si substrates. (**D**) Energy level diagram of the perovskite/C_{60} interface with MgF_x insertion layer.

with an x value in the range 1.0 ± 0.2. We expect this substoichiometric nature to produce a transverse electric dipole in this layer that promotes electron extraction.

To evaluate enhanced contact passivation with fluoride interlayers, we quantified the nonradiative recombination losses at the perovskite/ETL interface through absolute photoluminescence (PL) imaging under 1-sun equivalent illumination. This method let us extract the quasi–Fermi-level splitting (QFLS or Δμ) in the perovskite layer, which relates to the upper voltage limit of complete devices (*33*, *34*). Figure 2A and fig. S9 show that the mean QFLS of IZO/2PACz/perovskite structures without ETL was ~1.285 eV, whereas the IZO/2PACz/perovskite/C_{60} sample exhibited a sharp decline of QFLS with a mean value of 1.179 eV. The LiF- and MgF_x-treated samples displayed QFLS values of 1.198 and 1.217 eV, respectively. The HTL side remained unchanged, so we associated the undesired lower QFLS with trap-assisted recombination at the perovskite/ETL interface. Structural disorder or molecular imperfections in fullerene-based ETLs commonly have a strong band tail state (*35*, *36*) that may interact electronically with the perovskite layer to form undesired recombination channels.

Time-resolved photoluminescence spectroscopy further revealed that the IZO/2PACz/perovskite structure supported a very slow carrier decay process with an average carrier lifetime of ~1.6 μs (Fig. 2B and table S1). Coating C_{60} directly onto the perovskite expectedly caused a large reduction in PL lifetime to 26 ns, but the use of a MgF_x interlayer prolonged the average PL decay time to 83 ns, compared to 38 ns for the perovskite/C_{60} sample.

Transient absorption spectroscopy (TAS; see fig. S11) revealed a sharp negative band peaking at 718 and 710 nm for bare and C_{60}-coated perovskite samples, respectively, that could be assigned to ground-state photobleaching. As expected, the TAS signals of the perovskite/C_{60} sample exhibited faster decay of the bleaching peak than their perovskite/MgF_x/C_{60} counterparts. By globally fitting the TA decay curves of the three samples under four laser excitation conditions to a diffusion equation (fig. S10), we obtained a first-order charge-carrier decay constant k_1 of 3.48×10^5 s^{-1} ($1/k_1 \approx 2.87$ μs). This value was consistent with our PL decay and the electron-hole diffusion length of ~12 μm, which is much longer than the perovskite thickness, as desired for efficient solar cells.

These results indicate that the trap states causing nonradiative recombination mainly reside at the perovskite/ETL interface. To investigate the origin of these trap states, we conducted density functional theory (DFT) calculations with the structural model of the perovskite/C_{60} interface shown in Fig. 2C. The density of states (DOS) calculated at the relaxed contact distance showed the formation of deep trap states within the perovskite bandgap (Fig. 2C, inset). Similar calculations on prototypical $FAPbI_3$ perovskite confirmed the formation of such induced states (fig. S12). Notably, these states are not created by defects in the perovskite but are induced by proximity

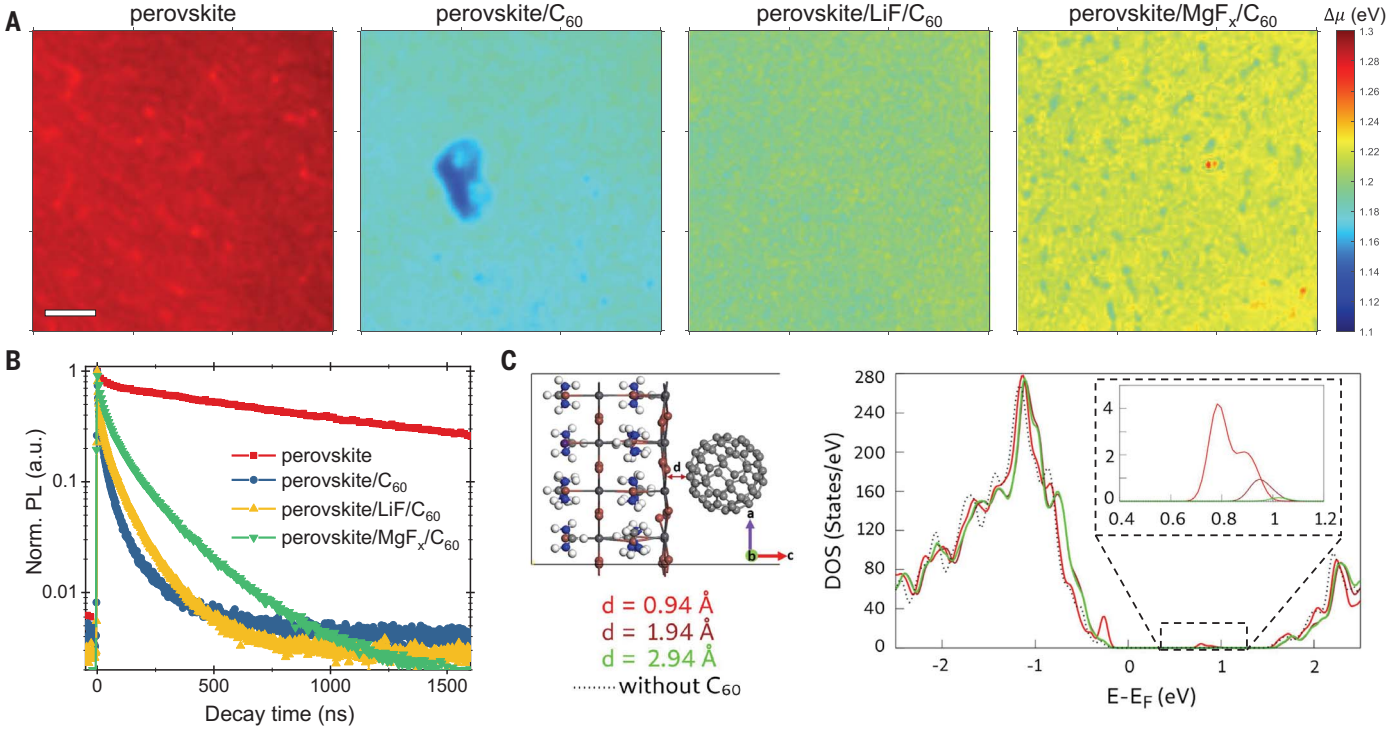


Fig. 2. Photoluminescence and transient absorption spectra. (**A**) Quasi–Fermi-level splitting (QFLS) mapping for the perovskite, perovskite/C_{60}, perovskite/LiF/C_{60}, and perovskite/MgF_x/C_{60} samples on a 2PACz-coated Si cell under 1-sun equivalent light intensity. Scale bar, 50 μm. (**B**) Time-resolved photoluminescence spectra of the bare perovskite and the perovskite films coated with different ETL structures. (**C**) Schematic model and DOS of a wide-bandgap perovskite as a function of the distance *d* to the C_{60} molecule. The inset of the enlarged DOS shows the induced midgap states when C_{60} is in close proximity to perovskite.

with C_{60} and are similar to metal-induced gap states in metal/semiconductor contacts (*37*, *38*). The MgF_x interlayer displaced C_{60} away from the perovskite and suppressed the induced trap states. On the basis of these findings, we argue that a key role of the interlayers is the blocking of the gap state–assisted recombination channels, thus suppressing charge recombination at the perovskite/ETL interface. In addition, DFT calculations showed that without interlayers, partial electron transfer from the perovskite into the C_{60} (fig. S16) created a barrier for electron extraction.

To verify improved charge extraction at the perovskite/C_{60} interface, we fabricated single-junction p-i-n devices with metal fluoride–based interlayers, as well as control samples without an interlayer. The solar cell with a MgF_x contact displacer reached a V_{oc} of 1.23 V, representing an absolute enhancement of ~50 mV relative to the control (fig. S17) and an enhancement of ~20 mV versus a LiF interlayer. These results agreed well with our energy-level and surface-passivation analyses. Notably, the FF improved to 81.1%, which we attributed to enhanced charge extraction and suppressed interface recombination at maximum power point conditions. In addition, we tested CaF_x devices and found that they also show remarkable V_{oc} and FF, implying that the alkali-earth metal fluoride as a contact displacer is a generic route to improve device performance.

We fabricated monolithic perovskite/silicon tandem solar cells using a double-textured Si bottom cell with a submicrometer random pyramid structure (fig. S19). The MgF_x-based device showed a remarkable reverse-scan PCE of up to 30.5% with a short-circuit current density J_{sc} of 19.8 mA/cm^2, V_{oc} of 1.92 V, and FF of 80.7% (Fig. 3A). The control tandem showed a best PCE of 28.6% with a J_{sc} of 19.8 mA/cm^2, V_{oc} of 1.85 V, and FF of 77.9% under reverse scan. The device statistics (Fig. 3, B and C, and fig. S20) corroborated that the PCE improvement was mainly the result of enhanced V_{oc} and FF. One unencapsulated MgF_x-based tandem was certified at Fraunhofer ISE CalLab, showing a reverse-scan PCE of 29.4% with a J_{sc} of 19.8 mA/cm^2, V_{oc} of 1.91 V, FF of 77.6%, and steady-state PCE of 29.3% (Fig. 3D and fig. S21). Integrating the calibrated external quantum efficiency (EQE; Fig. 3E) over the AM 1.5G spectrum yielded J_{sc} values of 20.0 and 19.8 mA/cm^2 for the perovskite and c-Si subcells, respectively, which agreed with our tandem J_{sc} values of ~19.8 mA/cm^2. Our optical analysis revealed that the optical loss, in addition to some reflection, mainly came from parasitic absorption in the IZO transparent top electrode and C_{60} layer (fig. S22), which accounted for equivalent values of 0.64 and 0.62 mA/cm^2, respectively.

To evaluate the perovskite subcell device performance, we conducted electroluminescence (EL) measurements on tandem devices. With an injected current density of 22 mA/cm^2, we observed well-resolved EL spectral mapping with peaks positioned at ~735 nm (Fig. 3, F and G), corresponding to the perovskite bandgap energy of ~1.69 eV. Under any current injection condition, the MgF_x-based tandem showed a relatively higher EL emission intensity than the control device, indicating a higher internal voltage (fig. S24). Combining the EL spectra results of the perovskite subcell under distinct current injection conditions with Suns-V_{oc} data of the c-Si single-junction cell, we constructed so-called pseudo–*J*-*V* curves of our perovskite/silicon tandem (Fig. 3H), free of any series resistance (R_s) losses. For the MgF_x-based tandem, we obtained a pseudo-V_{oc} of 1.93 V, which is remarkably near the V_{oc} of 1.92 V from standard *J*-*V* measurements. A pseudo-FF of 84.8% and pseudo-PCE of 32.5% could also be estimated, implying that ~3% in absolute PCE was lost to series resistance.

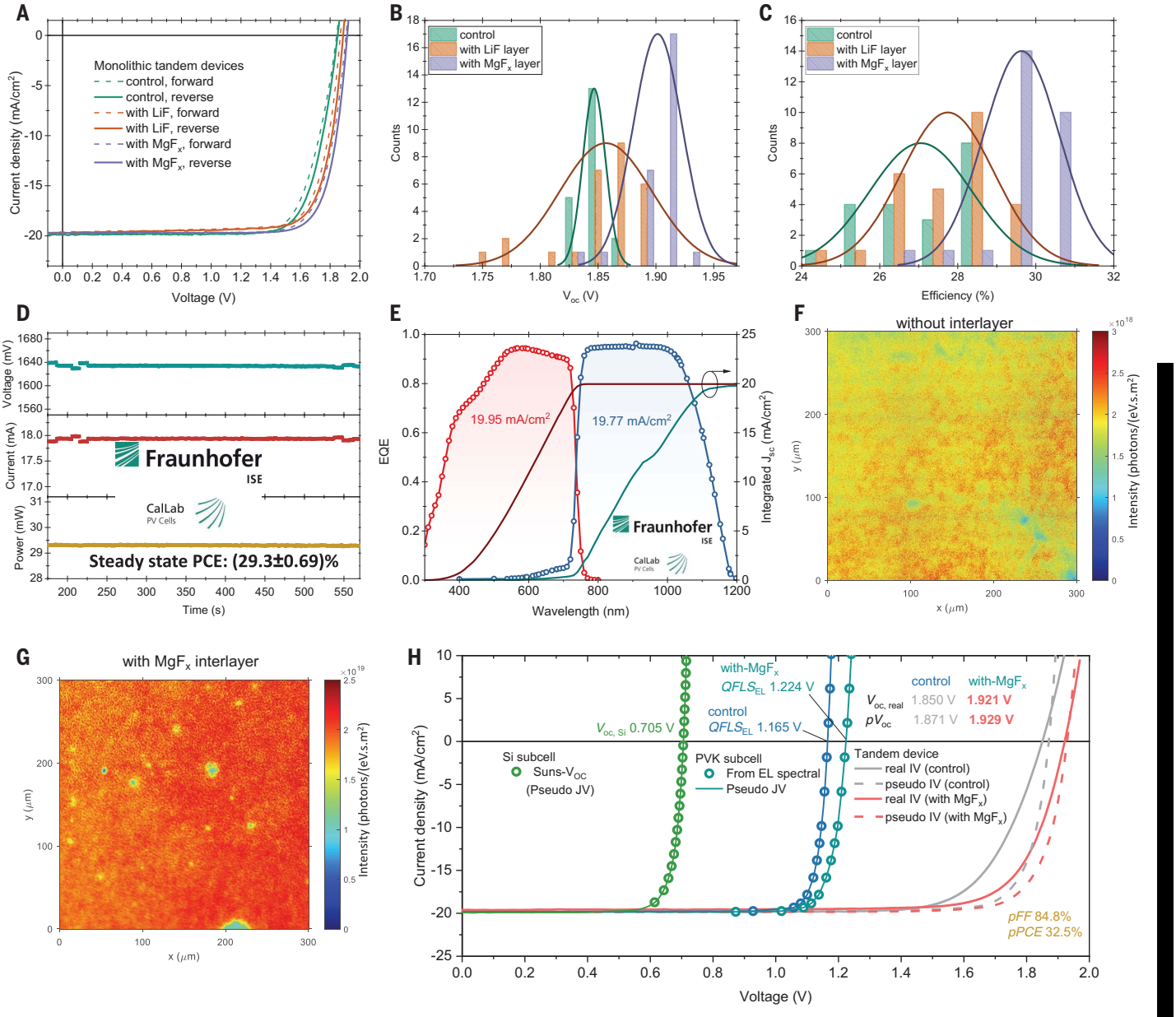


Fig. 3. Photovoltaic performance and pseudo–*J-V* characteristics. (**A**) *J-V* curves of the champion tandem cell. (**B** and **C**) Histogram of V_{oc} (B) and PCE (C) for the tandem solar cells fabricated in this study. (**D**) Stabilized power output of one MgF_x-based tandem device, certified by Fraunhofer ISE CalLab. (**E**) EQE spectra of the certified tandem device. (**F** and **G**) Absolute EL mapping of perovskite subcell without and with MgF_x interlayer under injection current of 22 mA/cm^2. (**H**) Reconstructed pseudo–*J-V* characteristics of our tandem device.

We explored the effect of the interlayer on device stability by monitoring the photovoltaic performance of the control and fluoride-based tandems without encapsulation under continual standard AM 1.5G illumination (Fig. 4A). The control device benefited from light soaking; its PCE increased from 27.2% initially to 28.0% after 10 min of illumination. The *J-V* curves (fig. S26) demonstrated that the light soaking improved V_{oc} and FF; we speculate that continuous illumination caused a slight favorable adjustment of the energy alignment at the perovskite/ETL interface. The fluoride-based devices did not appear to benefit from this (Fig. 4B and fig. S26), possibly because of the improved energy-level alignment at their perovskite/ETL interface. On a longer time scale, the LiF-based tandem, reputed to have low stability (*19*), showed a gradual performance drop from 29.1% to 27.5% in air, whereas the MgF_x-based tandem retained nearly >99% of its initial PCE after 260 min, which we attribute to MgF_x being nonhygroscopic and having a lower metal ion diffusivity. The control device maintained relatively stable but displayed lower absolute V_{oc} and FF values than the MgF_x-based devices after the light-soaking period.

In addition, we subjected our encapsulated tandem device to damp-heat testing (85°C with 85% relative humidity, IEC 61215:2021 standard; Fig. 4C and fig. S27). The MgF_x-treated tandem device did not show any V_{oc} or J_{sc} degradation after more than 1000 hours and retained 95.4% of its initial PCE. The V_{oc} even improved slightly, indicating that the perovskite itself and the interfacial layers were sufficiently tolerant to thermal stress. The FF showed a slight drop, which may be related

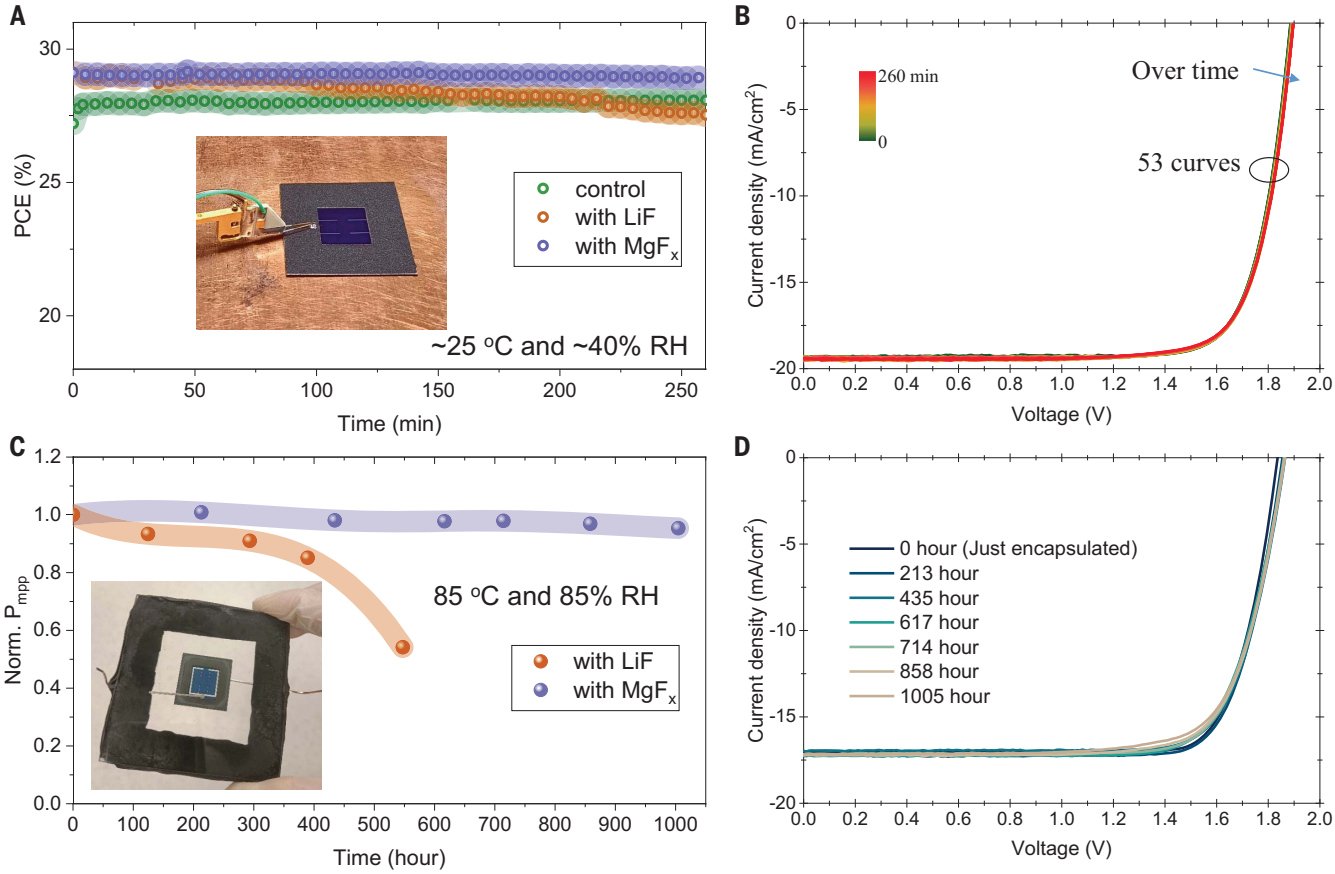


Fig. 4. Stability tests. (**A**) Evolution of the photovoltaic performance under continual AM 1.5G illumination for more than 4 hours in air. *J-V* scans were performed every ~5 min. The inset is the photo of one device under test with a black aperture mask on it. (**B**) *J-V* curves of the tandem devices with MgF_x interlayer under continual AM 1.5G illumination. (**C**) Evolution of photovoltaic performance of the encapsulated tandem devices when subjected to a damp-heat stability experiment at 85°C and 85% relative humidity. The inset is the photo of one encapsulated device. (**D**) *J-V* curves of the tandem devices during damp-heat tests.

to the increase in the series resistance of the contact electrode. In contrast, the LiF-treated tandem showed an obvious drop in PCE after an initial 125-hour testing (Fig. 4C and fig. S27). These results indicated that our MgF_x-tandem devices, with a reasonable encapsulation scheme, could pass the damp-heat test protocol of the IEC 61215:2021 standard.

REFERENCE AND NOTES

1. T. Leijtens, K. A. Bush, R. Prasanna, M. D. McGehee, *Nat. Energy* **3**, 828–838 (2018).
2. R. Wang *et al.*, *Nat. Photonics* **15**, 411–425 (2021).
3. H. Shen *et al.*, *Sci. Adv.* **4**, eaau9711 (2018).
4. E. Aydin *et al.*, *Energy Environ. Sci.* **14**, 4377–4390 (2021).
5. A. Richter, M. Hermle, S. W. Glunz, *IEEE Int. J. Photovolt.* **3**, 1184–1191 (2013).
6. B. Chen *et al.*, *Joule* **3**, 177–190 (2019).
7. J. P. Mailoa *et al.*, *Appl. Phys. Lett.* **106**, 121105 (2015).
8. B. Chen *et al.*, *Joule* **4**, 850–864 (2020).
9. J. Liu *et al.*, *Joule* **5**, 3169–3186 (2021).
10. M. Jošt *et al.*, *Energy Environ. Sci.* **11**, 3511–3523 (2018).
11. F. Sahli *et al.*, *Nat. Mater.* **17**, 820–826 (2018).
12. R. Santbergen *et al.*, *Opt. Express* **24**, A1288–A1299 (2016).
13. K. A. Bush *et al.*, *Nat. Energy* **2**, 17009 (2017).
14. D. Kim *et al.*, *Science* **368**, 155–160 (2020).
15. J. Xu *et al.*, *Science* **367**, 1097–1104 (2020).
16. Y. Hou *et al.*, *Science* **367**, 1135–1140 (2020).
17. S. Zhumagali *et al.*, *Adv. Energy Mater.* **11**, 2101662 (2021).
18. A. Al-Ashouri *et al.*, *Energy Environ. Sci.* **12**, 3356–3369 (2019).
19. A. Al-Ashouri *et al.*, *Science* **370**, 1300–1309 (2020).
20. E. L. Unger *et al.*, *J. Mater. Chem. A*11401–11409 (2017).
21. F. Peña-Camargo *et al.*, *ACS Energy Lett.* **5**, 2728–2736 (2020).
22. M. Stolterfoht *et al.*, *Energy Environ. Sci.* **12**, 2778–2788 (2019).
23. J. S. Luo *et al.*, *Energy Environ. Sci.* **11**, 2035–2045 (2018).
24. S.-G. Kim *et al.*, *Adv. Mater.* **33**, e2007431 (2021).
25. Z. Li *et al.*, *Energy Environ. Sci.* **10**, 1234–1242 (2017).
26. Z. Liu, J. Siekmann, B. Klingebiel, U. Rau, T. Kirchartz, *Adv. Energy Mater.* **11**, 2003386 (2021).
27. Q. Jiang *et al.*, *Nat. Photonics* **13**, 460– 466 (2019).
28. E. Aydin *et al.*, *Matter* **4**, 3549–3584 (2021).
29. J. Endres *et al.*, *J. Phys. Chem. Lett.* **7**, 2722–2729 (2016).
30. L. Canil *et al.*, *Energy Environ. Sci.* **14**, 1429–1438 (2021).
31. P. Schulz *et al.*, *Adv. Mater. Interfaces* **2**, 1400532 (2015).
32. Y. Wan *et al.*, *ACS Appl. Mater. Interfaces* **8**, 14671–14677 (2016).
33. D. W. deQuilettes *et al.*, *ACS Energy Lett.* **1**, 438–444 (2016).
34. M. Stolterfoht *et al.*, *Adv. Mater.* **32**, 2000080 (2020).
35. F. Bussolotti *et al.*, *Phys. Rev. B* **92**, 115102 (2015).
36. Y. Shao, Y. Yuan, J. Huang, *Nat. Energy* **1**, 15001 (2016).
37. M. Sajjad *et al.*, *Appl. Phys. Lett.* **114**, 071601 (2019).
38. J. Robertson, *J. Vac. Sci. Technol. B* **18**, 1785–1791 (2000).

ACKNOWLEDGMENTS

We acknowledge the use of KAUST Solar Center and Core Lab facilities and the support of its staff. **Funding:** Supported by the King Abdullah University of Science and Technology (KAUST) under awards OSR-CARF/CCF-3079, OSR-CRG2019-4093, OSR-CRG2020-4350, IED OSR-2019-4208, IED OSR-2019-4580, REI/1/4833-01-01, and OSR-CRG2018-3737. **Author contributions:** Conceptualization: J.L., E.A., S.D.W.; methodology: J.L., M.D.B., Y.G., G.T.H., R.R.P., M.K.E., M.M., A.S., E.A., L.X.; investigation: J.L., M.D.B., Y.G., G.T.H., R.R.P., M.K.E., M.M., A.S., W.Y., M.B., F.X., L.X., Au.R., A.R., T.G.A., M.W., J.K., A.A.S.; visualization: J.L., G.T.H., R.R.P., M.K.E., M.M., A.S., Y.G., A.S.S., E.U., M.W.; funding acquisition: S.D.W., F.L., D.A., U.S.; supervision: S.D.W., F.L., D.A., U.S.; writing–original draft: J.L., E.A., S.D.W.; writing–review and editing: J.L., S.D.W., E.A., E.U., M.D.B., F.L., Y.G., G.T.H., F.H.I., R.A. **Competing interests:** J.L. and S.D.W. are inventors on patent application (63/312,896) submitted by King Abdullah University of Science and Technology that covers the alkali-earth metal fluoride used in this work. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn8910
Materials and Methods
Supplementary Text
Figs. S1 to S27
Table S1
References (*39*–*56*)

Submitted 28 December 2021; resubmitted 25 March 2022
Accepted 13 June 2022
Published online 23 June 2022
10.1126/science.abn8910

SOLAR CELLS

Accelerated aging of all-inorganic, interface-stabilized perovskite solar cells

Xiaoming Zhao[1], Tianran Liu[1,2], Quinn C. Burlingame[3], Tianjun Liu[4], Rudolph HolleyIII[1], Guangming Cheng[5], Nan Yao[5], Feng Gao[4], Yueh-Lin Loo[1]*

To understand degradation routes and improve the stability of perovskite solar cells (PSCs), accelerated aging tests are needed. Here, we use elevated temperatures (up to 110°C) to quantify the accelerated degradation of encapsulated $CsPbI_3$ PSCs under constant illumination. Incorporating a two-dimensional (2D) $Cs_2PbI_2Cl_2$ capping layer between the perovskite active layer and hole-transport layer stabilizes the interface while increasing power conversion efficiency of the all-inorganic PSCs from 14.9 to 17.4%. Devices with this 2D capping layer did not degrade at 35°C and required >2100 hours at 110°C under constant illumination to degrade by 20% of their initial efficiency. Degradation acceleration factors based on the observed Arrhenius temperature dependence predict intrinsic lifetimes of 51,000 ± 7000 hours (>5 years) operating continuously at 35°C.

Although the power-conversion efficiencies (PCEs) of metal halide perovskite solar cells (PSCs) can now exceed 25%, long-term operational instability issues must be addressed before they can be commercialized (*1*–*3*). The most stable and efficient PSCs have reported T_{80} lifetimes (the time at which the PCE drops to 80% of its initial value) of just a few hundred or thousand hours (*4*–*7*) under continuous illumination, versus the >20-year lifetimes required for most commercial applications. Despite an encouraging report of PSCs with a T_{80} of more than 1 year, the PCE of the solar cells is relatively low (<13%) (*8*). Accelerated aging tests can facilitate rapid PSC stability screening (*9*–*11*). An effective accelerated stress test can quantify the lifetime acceleration factor (AF) that relates the lifetime under a defined standard operating condition to the lifetime under elevated stress conditions. Thermal and high-intensity illumination stress tests with AFs have been developed for organic- and silicon-based PVs (*11*–*15*), but developing robust AFs for PSCs is challenging given their complex sensitivities to temperature, light, and electrical bias (*10*, *16*, *17*).

Here, we thermally accelerated the degradation of encapsulated $CsPbI_3$ PSCs operating under constant illumination at their maximum power point (MPP) with temperatures from 35° to 110°C. The PCE degradation rate followed an Arrhenius temperature dependence that was slowed substantially by incorporating a two-dimensional (2D) $Cs_2PbI_2Cl_2$ capping layer between the perovskite active layer and the hole transport layer (HTL). This 2D layer stabilized the perovskite/HTL interface and suppressed ion migration into the HTL, resulting in a T_{80} lifetime of >2100 hours at 110°C. With an experimentally determined AF of 24.2 ± 3.5, this lifetime corresponds to an extrapolated T_{80} of $5.1 \pm 0.7 \times 10^4$ hours—or more than 5 years of continuous operation at 35°C.

The most efficient PSCs have been based on hybrid organic-inorganic perovskites, which contain volatile organic cations, such as methylammonium (MA^+) and formamidinium (FA^+). To maximize the thermal stability and photostability of our solar cells, we chose inorganic $CsPbI_3$ as the photoabsorber (*18*–*20*), despite its solar cells exhibiting slightly lower efficiencies. Inorganic $CsPbI_3$ PSCs with the structure shown in Fig. 1A were fabricated both with (capped) and without (uncapped) a 2D $Cs_2PbI_2Cl_2$ layer between the $CsPbI_3$ absorber and the CuSCN HTL. The all-inorganic stack that includes TiO_2, Al_2O_3, and CuSCN transport layers, as well as fluorinated tin oxide (FTO) and Cr/Au electrodes, was designed to maximize the thermal stability and photostability of these devices, as discussed in supplementary text 1 and fig. S1. Photovoltaic characterization of representative as-fabricated devices is shown in Fig. 1, B and C, and fig. S2 and listed in table S1, with population statistics as shown in fig. S3. Capped PSCs have improved fill factors (FFs) and open-circuit voltages (V_{OC}'s), leading to a champion PCE of 17.4% compared with 14.9% for uncapped devices. This PCE is

[1]Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA. [2]Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA. [3]Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ 08544, USA. [4]Department of Physics, Chemistry, and Biology (IFM), Linköping University, SE-58183 Linköping, Sweden. [5]Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, NJ 08544, USA.
*Corresponding author. Email: lloo@princeton.edu

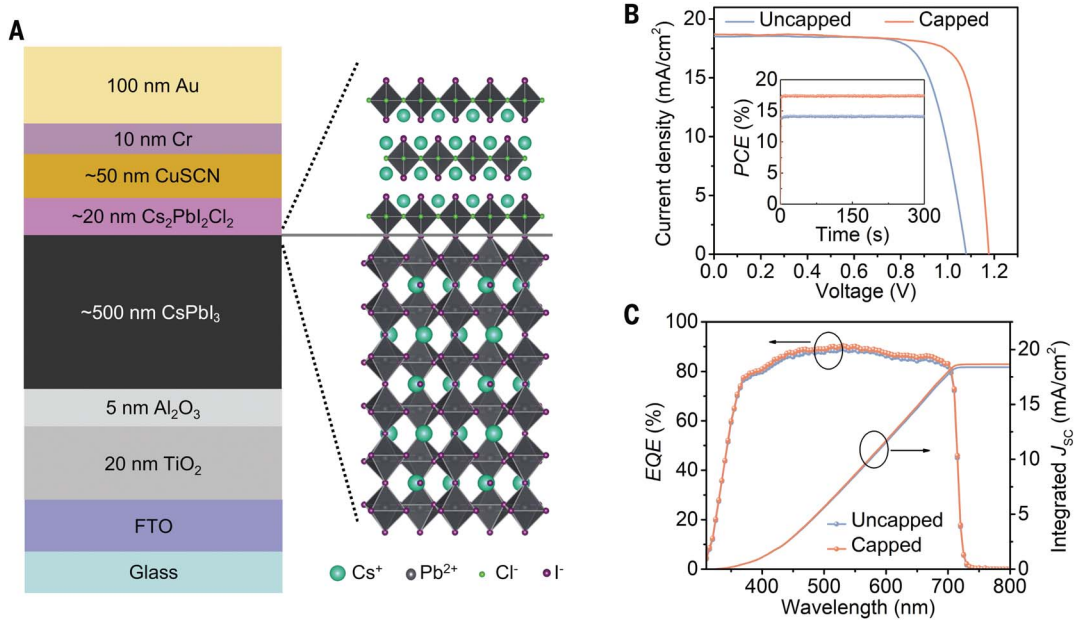


Fig. 1. Device structure and photovoltaic characterization. (**A**) Illustration of the capped device structure used in this work with a 2D $Cs_2PbI_2Cl_2$ layer atop the 3D perovskite active layer. (**B**) Current-density versus voltage (*J*-*V*) characteristics under reverse voltage scan (inset: stabilized power output) and (**C**) external quantum efficiency (EQE) spectra and integrated J_{SC} of capped and uncapped champion solar cells.

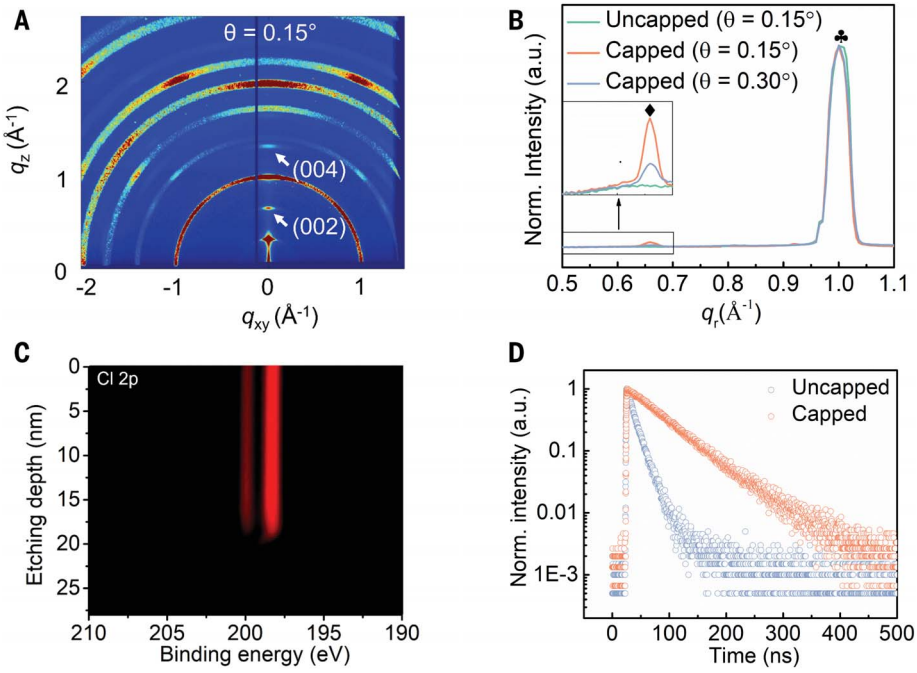


Fig. 2. Characterization of 2D perovskite capping layers. (**A**) GIWAXS patterns of a $Cs_2PbI_2Cl_2/CsPbI_3$ perovskite stack taken at a surface-sensitive incident angle of 0.15°. Characteristic reflections of $Cs_2PbI_2Cl_2$ are indicated. (**B**) Azimuthally integrated GIWAXS traces collected on uncapped and capped samples at incident angles of θ = 0.15° (more surface-sensitive) and θ = 0.30° (less surface-sensitive). For reference, the critical angle is θ = 0.24°. The diamond indicates the (002) reflection of $Cs_2PbCl_2I_2$ and the clover indicates the (110) reflection of $CsPbI_3$. (**C**) XPS Cl 2p depth profiling of the $Cs_2PbI_2Cl_2/CsPbI_3$ stack. (**D**) Transient photoluminescence of capped and uncapped films on glass, excited from the film side.

the highest among fully inorganic PSCs in which all the functional materials in the stack are inorganic.

Depositing a thin 2D perovskite capping layer can stabilize the surfaces of organic-inorganic hybrid perovskites, such as $FAPbI_3$ and $MAPbI_3$ (*21*), but this approach has not yet been successfully implemented with inorganic perovskites. This challenge stems from the stronger binding strength of Cs^+ compared to MA^+ or FA^+, which prevents cation exchange between the Cs^+ of $CsPbI_3$ and the organic ligands in the applied hybrid organic-inorganic 2D perovskite precursor solution (*22*–*25*). To avoid this problem, we deposited a fully inorganic $Cs_2PbI_2Cl_2$ 2D layer by treating the $CsPbI_3$ surface with a CsCl solution followed by thermal annealing. Grazing-incidence wide-angle x-ray scattering (GIWAXS) patterns (Fig. 2A and fig. S4) showed two new reflections emerging on $CsPbI_3$ films after CsCl treatment corresponding to the (002) and (004) reflections of 2D $Cs_2PbI_2Cl_2$. Increasing the incident angle of the x-ray beam resulted in a decrease in the relative intensity of these reflections (Fig. 2B), suggesting that the 2D layer formed preferentially on the $CsPbI_3$ surface. The interfacial nature of this 2D layer was also confirmed by cross-sectional scanning electron microscopy (SEM) imaging of the device (fig. S5). We estimated the thickness of the capping layer to be 20 nm by tracking chlorine content on the film surface with x-ray photoelectron spectroscopy (XPS) while depth etching (Fig. 2C).

To study the surface passivation effect of this capping layer, we measured time-resolved photoluminescence (TRPL) transients (Fig. 2D). The lifetime of the device increased from 14 to >62 ns in the presence of the capping layer, suggesting that it effectively suppressed nonradiative recombination at the $CsPbI_3$ surface and extended the lifetime and diffusion length of charge carriers (*26*). This observation is consistent with the observed V_{OC} increase in capped PSCs (Fig. 1B), whose PV characteristics are tabulated in table S1. This trend was also comparable to previous reports of 2D and 3D PSCs based on organic-inorganic hybrid perovskites that benefited from surface passivation (*26*–*28*).

To evaluate the stability of capped and uncapped PSCs, N_2-encapsulated solar cells (fig. S6) with both structures were aged at their MPP under illumination from a metal halide solar simulator (spectrum shown in fig. S7) at 35°, 59°, 85°, and 110°C. Our stability studies used solar cells having the same device configuration as those for PV characterization, including encapsulation and the light aperture. The evolution of the normalized PCE (averaged from three subcells) for both device types at these operating temperatures, and the evolution for the unnormalized data, normalized short-circuit current (J_{SC}), V_{OC}, FF, and external quantum efficiency (EQE), are shown in figs. S8 to S10. Figure S11 shows the operational stability of 2D and 3D PSCs without encapsulation operating at 110°C, confirming the robustness of our solar cell packaging. Heating the solar cells reversibly reduced their PCEs (*29*), which accounts for the differences in the initial PCE of the unnormalized data set in fig. S8. We observe a clear temperature dependence of the PCE degradation rate that can be fitted to a biexponential function of the form

$$PCE(t) = A1{*}\exp(-k_{fast} \times t) + A2{*}\exp(-k_{slow} \times t) + B \quad (1)$$

with a fast and slow degradation rate, k_{fast} and k_{slow}. Here, $A1$, $A2$, and B are constants and t is time. We fit the data collected across the temperature range using this function with an $R^2 > 0.95$, except for the data acquired on the capped solar cells operating at 35°C because no PCE degradation was observed, even after 3531 hours of continuous operation (Fig. 3B).

To estimate an activation energy for the degradation processes corresponding to k_{fast} and k_{slow}, we assume an Arrhenius temperature dependence to the degradation rates

$$k(T) = A\exp\left(\frac{-E_a}{k_B T}\right) \quad (2)$$

where $k(T)$ is a degradation rate at temperature T, A is constant, E_a is the activation energy of degradation, and k_B is Boltzmann's constant (*11*–*15*). Rearranging Eq. 2, the activation energy is equivalent to the slope:

$$E_a = -\frac{\partial \ln(k(T))}{\partial\left(\frac{1}{k_B T}\right)} \quad (3)$$

The temperature-dependent degradation rates were plotted as a function of inverse temperature in Fig. 3C and fitted to Eq. 3. Degradation rates could be adequately described by a single Arrhenius function across the entire temperature range. This finding suggested that the same degradation mechanism dominated the entire temperature range, which is an important criterion for a reliable accelerated aging test (*9*). We observed that the activation energies associated with the fast and slow degradation of each type of PSC were comparable, suggesting that the two degradation rates probed a single physical process. As we will discuss in detail below, we speculate that ion migration is the dominant degradation mechanism. The E_a's that describe the degradation for capped PSCs are nearly twice those for uncapped PSCs, suggesting that the 2D $Cs_2PbI_2Cl_2$ layer stabilizes the devices against thermal degradation stemming from ion migration.

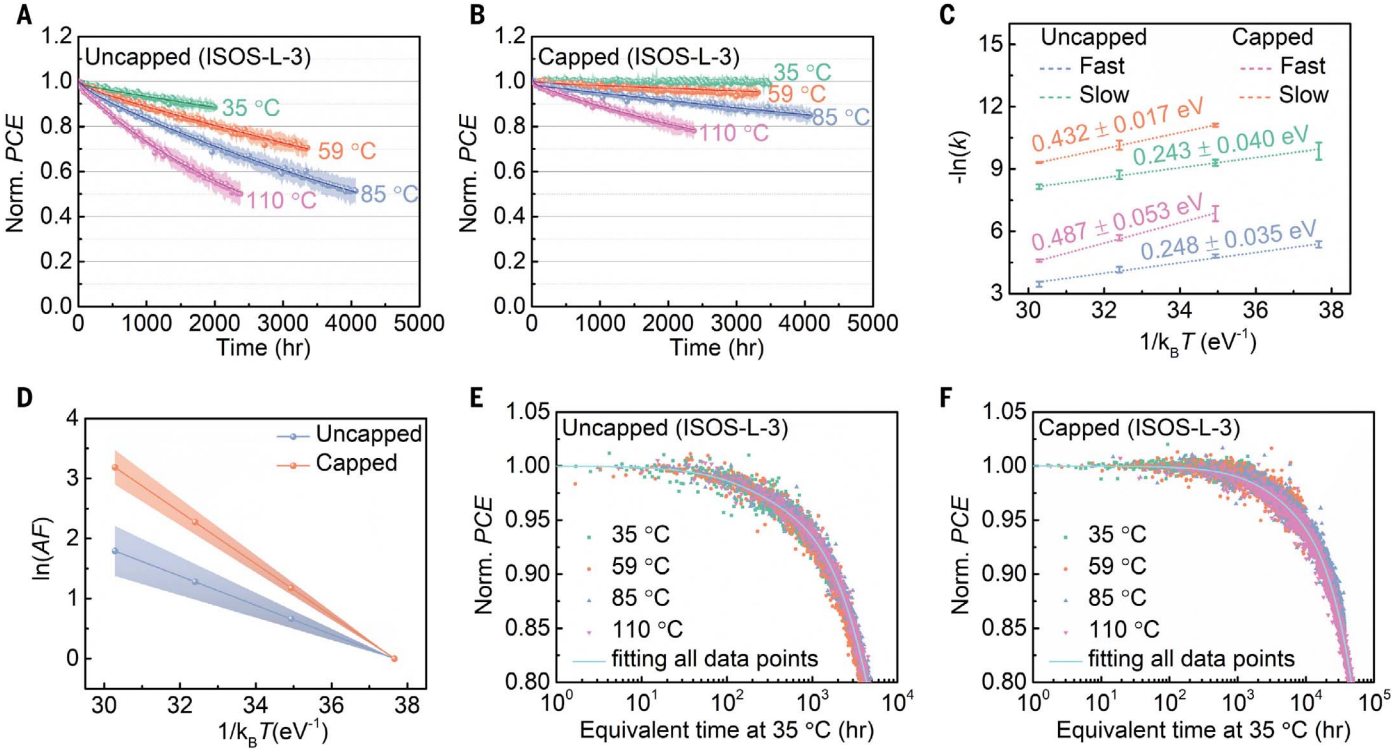


Fig. 3. Accelerated aging of PSCs operating at elevated temperatures. Operational stability of (**A**) uncapped and (**B**) capped PSCs operating at 35°, 59°, 85°, and 110°C, with standard deviation envelopes. The encapsulated devices were held at their MPP under constant full-spectrum simulated sunlight (power density ≈ 120 mW/cm^2) in (65 ± 26)% relative humidity air. Data points represent average PCE of three devices fabricated and tested under the same condition. The solid lines are biexponential fits to the data. (**C**) Natural logarithm of degradation rates (k_{fast} and k_{slow}) versus $1/k_BT$ obtained from biexponential fits to PCEs for uncapped and capped PSCs, where T is the aging temperature. Linear fits used to extract the activation energy (E_a) from each exponential are shown as the dash lines. Because the capped solar cells operating at 35°C did not show noticeable degradation over the time interval studied, their fitted values were not included. (**D**) Natural logarithm of AF versus $1/k_BT$ with the standard deviation represented by the shaded area around each line. Standard operating conditions for the AF calculation were defined as 1 sun illumination and T_{ref} = 35°C. (**E** and **F**) Normalized PCE of uncapped and capped PSCs plotted against the equivalent aging time at 35°C, defined as aging time (in units of hours) multiplied by the acceleration factor, for PSCs under all temperatures studied in this work.

To extract a meaningful degradation rate at standard operating conditions, k_{ref}, from the accelerated degradation rate at high temperatures, k_{acc}, we define an AF (*11*):

$$AF = \frac{k_{acc}}{k_{ref}} = \exp\left(\frac{E_a}{k_B}\left[\frac{1}{T_{ref}} - \frac{1}{T_{acc}}\right]\right) \quad (4)$$

where T_{acc} and T_{ref} are the operating temperatures during aging at accelerated and standard operating conditions (defined as 1 sun intensity, 35°C). Because k_{slow} is primarily responsible for long-term degradation and the E_a's of the two rates are nearly identical, we used the E_a extracted from k_{slow} and Eq. 4 to obtain AFs for each T_{acc} (specifically 59°, 85°, and 110°C) (Fig. 3D). On the basis of these AFs, we can define an equivalent operating time at T_{ref} = 35°C for devices aged at elevated temperatures by multiplying their aging time by the acceleration factor (Fig. 3, E and F). When processed in this manner, all of the data collapsed onto a universal curve for both capped and uncapped solar cells, which further confirmed that the same mechanism, albeit hastened at elevated temperatures, was responsible for degradation across the entire temperature range. With this dataset, we can make T_{80} lifetime predictions for capped devices operating at 35°C. This estimate is critical because devices aged at these conditions do not show any PCE degradation, even after continuous operation for 3531 hours. Given that the average experimentally measured T_{80} for capped solar cells operating continuously at 110°C is 2100 hours (Fig. 3B), and the AF for the capped solar cells at 110°C is 24.2 ± 3.5, we estimated a T_{80} of 5.1 ± 0.7 × 10^4 hours at 35°C.

To investigate the origins of PSC degradation, we aged full-stack devices at 110°C under continuous illumination before peeling off their top electrodes and characterizing the structure of the active layers with x-ray diffraction (XRD). As shown in Fig. 4, A and B, the reflection at 2θ = 16.15° corresponding to the (003) reflection of CuSCN in the XRD trace of the uncapped PSCs became broader and less intense after 2000 hours of aging, which suggested a decrease in the CuSCN crystallite size. Clear degradation of the CuSCN surface in aged, uncapped PSCs was also visible in the SEM images in fig. S12, which show the formation of pinholes and decreased film uniformity. Correspondingly, the I 3d signal increased substantially in the XPS spectrum of the aged HTL surface (Fig. 4C). These results suggest that iodine migration from the $CsPbI_3$ active layer was responsible for the structural changes in CuSCN in uncapped solar cells, as discussed in supplementary text 2 and figs. S13 and S14. By contrast, the film structure and morphology of CuSCN in capped PSCs did not show appreciable changes after the same duration, as shown in their XRD patterns and SEM images (Fig. 4, A and B, and fig. S12), and we did not observe any appreciable I 3d signal in the XPS spectrum of the HTL surface of capped PSCs (Fig. 4C).

To quantify the ease of ion migration in capped and uncapped films, we measured the temperature-dependent conductivity in two-terminal lateral devices, a widely used method for characterizing ion migration in perovskite films (*4*, *30*, *31*). The activation

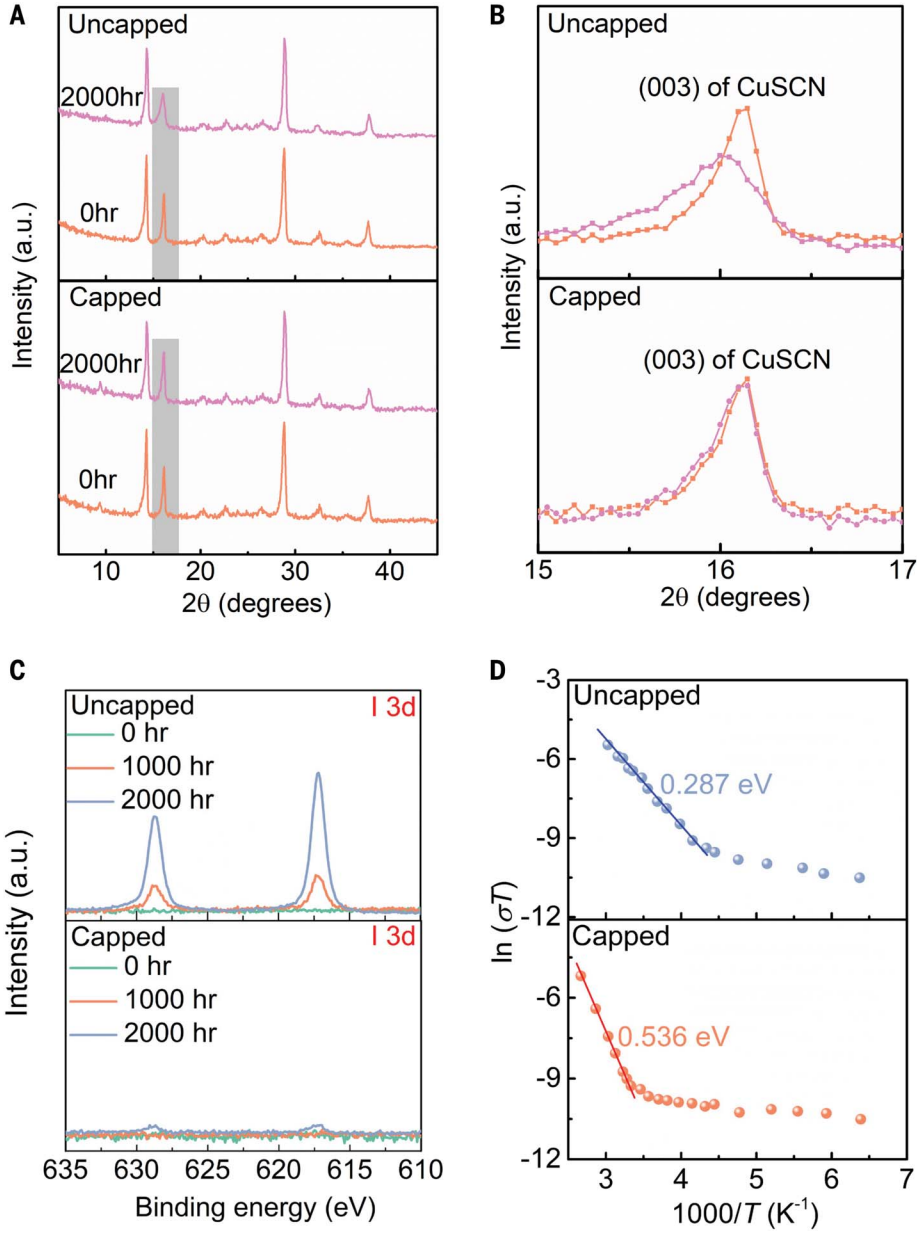


Fig. 4. Degradation and stabilization mechanism. (**A**) XRD traces of uncapped and capped PSCs before and after aging for 2000 hours, with (**B**) the region corresponding to CuSCN (003) diffraction magnified. (**C**) XPS I 3d spectra of the CuSCN surface in uncapped or capped PSCs with aging time on removal of the top electrode. For studies in (A) to (C), full-stack, functional devices were sealed in N_2 and placed under full-sun illumination at 110°C for the specified time before the electrodes were peeled off prior to measurement. (**D**) Temperature-dependent conductivity (σ) of uncapped and capped films.

energy of ion migration ($E_{a,ion}$) values was extracted by fitting the measured temperature-dependent conductivity (σ) to the Nernst-Einstein equation

$$\sigma(T) = \frac{\sigma_0}{T}\exp\left(\frac{-E_{a,ion}}{k_B T}\right) \quad (5)$$

where σ_0 is a temperature-independent prefactor (*30*, *31*). Plotting $1/T$ versus $\ln(\sigma T)$, two distinct regions were present in both capped and uncapped films corresponding to two distinct transport regimes. The high-temperature linear regime, corresponding to ion-dominated transport, was used to extract an approximate $E_{a,ion}$ (Fig. 4D). Similar to the activation of degradation, the $E_{a,ion}$ of the capped film is nearly twice that of the uncapped film, indicating that the 2D cap frustrated ion migration. Suppressed ion migration likely stems from passivation of iodine vacancies at the surface of the perovskite active layer in the presence of a 2D capping layer (*4*, *32*, *33*), as shown in fig. S15.

REFERENCES AND NOTES

1. M. Grätzel, *Nat. Mater.* **13**, 838–842 (2014).
2. N. Li, X. Niu, Q. Chen, H. Zhou, *Chem. Soc. Rev.* **49**, 8235–8286 (2020).
3. H. Min *et al.*, *Nature* **598**, 444–450 (2021).
4. S. Yang *et al.*, *Science* **365**, 473–478 (2019).
5. Z. Dai *et al.*, *Science* **372**, 618–622 (2021).
6. N. Arora *et al.*, *Science* **358**, 768–771 (2017).
7. Y. H. Lin *et al.*, *Science* **369**, 96–102 (2020).
8. G. Grancini *et al.*, *Nat. Commun.* **8**, 15684 (2017).
9. M. V. Khenkin *et al.*, *Nat. Energy* **5**, 35–49 (2020).
10. K. M. Anoop *et al.*, *Sol. RRL* **4**, 1900335 (2020).
11. O. Haillant, D. Dumbleton, A. Zielnik, *Sol. Energy Mater. Sol. Cells* **95**, 1889–1895 (2011).
12. J. Kettle *et al.*, *Sol. Energy Mater. Sol. Cells* **167**, 53–59 (2017).
13. Q. Burlingame *et al.*, *Adv. Energy Mater.* **6**, 1601094 (2016).
14. S. Züfle, R. Hansson, E. A. Katz, E. Moons, *Sol. Energy* **183**, 234–239 (2019).
15. Q. Burlingame *et al.*, *Nature* **573**, 394–397 (2019).
16. Z. Wang *et al.*, *Nat. Energy* **3**, 855–861 (2018).
17. C. Law *et al.*, *Adv. Mater.* **26**, 6268–6273 (2014).
18. S. M. Yoon *et al.*, *Joule* **5**, 183–196 (2021).
19. Y. Wang *et al.*, *Angew. Chem. Int. Ed.* **58**, 16691–16696 (2019).
20. Y. Wang *et al.*, *Science* **365**, 591–595 (2019).
21. D. L. McGott *et al.*, *Joule* **5**, 1057–1073 (2021).
22. X. Liu *et al.*, *Angew. Chem. Int. Ed.* **60**, 12351–12355 (2021).
23. Y. Wang *et al.*, *Joule* **2**, 2065–2075 (2018).
24. E. M. Sanehira *et al.*, *Sci. Adv.* **3**, eaao4204 (2017).
25. Y. Wang, T. Zhang, M. Kan, Y. Zhao, *J. Am. Chem. Soc.* **140**, 12345–12348 (2018).
26. Q. Jiang *et al.*, *Nat. Photonics* **13**, 460–466 (2019).
27. X. Jiang *et al.*, *ACS Appl. Mater. Interfaces* **13**, 2558–2565 (2021).
28. W. Shi, H. Ye, *J. Phys. Chem. Lett.* **12**, 4052–4058 (2021).
29. T. Moot *et al.*, *ACS Energy Lett.* **6**, 2038–2047 (2021).
30. J. Xing *et al.*, *Phys. Chem. Chem. Phys.* **18**, 30484–30490 (2016).
31. X. Li *et al.*, *Nat. Commun.* **9**, 3806 (2018).
32. B. Chen, P. N. Rudd, S. Yang, Y. Yuan, J. Huang, *Chem. Soc. Rev.* **48**, 3842–3867 (2019).
33. X. Zhao, T. Liu, Y.-L. Loo, *Adv. Mater.* **34**, e2105849 (2022).

ACKNOWLEDGMENTS

The authors thank E. Tsai and R. Li from Brookhaven National Laboratory for their assistance with x-ray scattering measurements. **Funding:** X-ray scattering measurements were conducted at the Center for Functional Nanomaterials (CFN) and the Complex Materials Scattering (CMS) beamline of the National Synchrotron Light Source II (NSLS-II), which both are US DOE Office of Science Facilities, at Brookhaven National Laboratory under Contract No. DE-SC0012704. The authors acknowledge the use of Princeton's Imaging and Analysis Center, which is partially supported through the Princeton Center for Complex Materials (PCCM), a National Science Foundation (NSF)-MRSEC program (DMR-2011750). Y.-L.L. acknowledges support from the National Science Foundation, under grants DMR-1627453 and CMMI-1824674. F.G. acknowledges the Swedish Government Strategic Research Area in Materials Science on Functional Materials at Linköping University (Faculty Grant SFO-Mat-LiU No. 2009-00971). T.R.L. thanks the Julis Romo Rabinowitz Graduate Fellowship for funding. Q.C.B. thanks the Arnold and Mabel Beckman Foundation for supporting this work. **Author contributions:** Y.-L.L. directed and supervised the project. X.Z. conceived the idea, initialized the project, and contributed to films and devices fabrication and characterization. T.R.L. and Q.C.B. contributed to accelerated stability studies. T.J.L. and F.G. contributed to temperature-dependent conductivity studies. R.H. contributed to XRD refinements. G.C. and N.Y. contributed to device stack imaging. All authors contributed to discussions and to finalizing the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn5679
Materials and Methods
Supplementary Text
Figs. S1 to S15
Table S1
References (*34*–*55*)

Submitted 5 December 2021; accepted 19 April 2022
Published online 16 June 2022
10.1126/science.abn5679

QUANTUM SIMULATION

Thermalization dynamics of a gauge theory on a quantum simulator

Zhao-Yu Zhou[1,2,3,4]*†, Guo-Xian Su[1,2,3,4]†, Jad C. Halimeh[5], Robert Ott[6], Hui Sun[1,2,3,4], Philipp Hauke[5], Bing Yang[3,7]‡, Zhen-Sheng Yuan[1,2,3,4,8], Jürgen Berges[6], Jian-Wei Pan[1,2,3,4,8]

Gauge theories form the foundation of modern physics, with applications ranging from elementary particle physics and early-universe cosmology to condensed matter systems. We perform quantum simulations of the unitary dynamics of a U(1) symmetric gauge field theory and demonstrate emergent irreversible behavior. The highly constrained gauge theory dynamics are encoded in a one-dimensional Bose-Hubbard simulator, which couples fermionic matter fields through dynamical gauge fields. We investigated global quantum quenches and the equilibration to a steady state well approximated by a thermal ensemble. Our work may enable the investigation of elusive phenomena, such as Schwinger pair production and string breaking, and paves the way for simulating more complex, higher-dimensional gauge theories on quantum synthetic matter devices.

Gauge theories provide a fundamental description of quantum dynamics in the Standard Model of particle physics. However, unitary quantum evolution admits no loss of information on a fundamental level, so describing from first principles the emergence of phenomena such as thermalization in gauge theories is an outstanding challenge in physics. No general method exists that can simulate the time evolution of the underlying complex quantum many-body theory on classical computers (*1*, *2*). Much progress on emergent phenomena has been achieved for simpler systems (*3*, *4*). For gauge fields, however, the direct connection of far-from-equilibrium behavior at early evolution times with the possible late-time approach to thermal equilibrium, for example, as indicated in collisions of heavy nuclei, remains elusive (*5*).

Quantum simulators open up a way forward to address this long-standing problem. In recent years, there has been much progress in the engineering of gauge theories using various quantum resources such as trapped ions (*6*, *7*), cold atomic gases (*8*–*12*), arrays of Rydberg atoms (*13*, *14*), and superconducting qubits (*15*, *16*). Such tabletop platforms can give access to a plethora of observables with high resolution in time and space. However, the simulation requires a large-scale system to incorporate the many degrees of freedom required for the complex quantum field dynamics. In addition, because gauge theories are governed by local symmetries, the engineering of the many gauge constraints at each point in space and time during a nonequilibrium evolution provides a major challenge.

Here, we performed quantum simulations of the far-from-equilibrium dynamics of a U(1) symmetric gauge field theory and demonstrate the emergence of thermal equilibrium properties at late times. To achieve this, we used a large-scale Bose-Hubbard quantum simulator (*8*) and precisely controlled the highly excited states relevant for the nonequilibrium dynamics of the gauge theory. The system couples fermionic matter through dynamical gauge fields in one spatial dimension and uses a discrete "quantum-link" (*17*) representation, which is also discussed in condensed matter physics (*18*, *19*) and in the context of particle physics (*20*, *21*).

Exploiting the full experimental tunability of the Bose-Hubbard model parameters, we explored the influence of the gauge-symmetry constraints on the evolution and established the thermalization dynamics of the U(1) gauge theory. After quenching from gauge-invariant initial states far from equilibrium, we observed emergent many-body oscillations through the dynamical annihilation and creation of fermion pairs. We demonstrated an effective loss of information about the system's initial state by starting from different initial conditions with the same conserved quantities and observing relaxation toward a common steady state at longer times. These thermalization dynamics are illustrated in Fig. 1A.

The unitary dynamics of the one-dimensional lattice gauge theory is governed by the Hamiltonian

$$\hat{H}_{\text{gauge}} = \sum_l \left[\frac{\kappa}{2} \left(\hat{\psi}_l \hat{S}^+_{l,l+1} \hat{\psi}_{l+1} + \text{H.c.} \right) + m \hat{\psi}^\dagger_l \hat{\psi}_l \right]$$

where $\hat{\psi}_l^{(\dagger)}$ are fermionic field operators (*22*) on matter site l with mass m; $\hat{S}^{+(-)}_{l,l+1}$ are spin-1/2 raising (lowering) operators for the gauge fields on the link between matter sites l and $l+1$; and interaction ~κ represents the annihilation (or creation) of a pair of fermionic charges on neighboring sites with a concomitant change of electric flux $\hat{E}_{l,l+1} = (-1)^{l+1} \hat{S}^z_{l,l+1}$ on the gauge link in between, such that gauge invariance is retained. The model is realized within a subspace of our quantum simulator, which is described by a tilted Bose-Hubbard Hamiltonian with a staggered potential (see Eq. S5 for details). It is characterized by direct

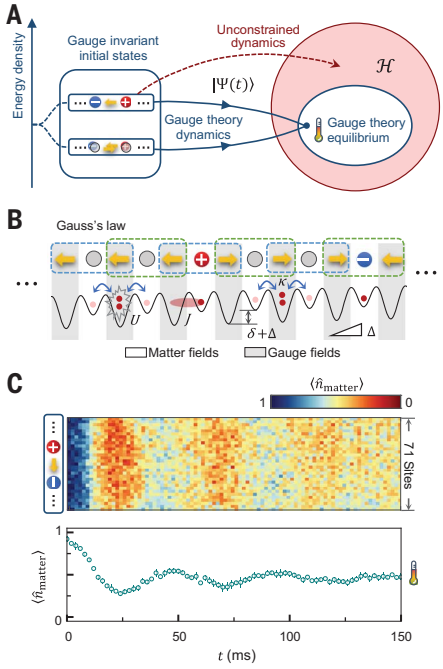


Fig. 1. Quantum simulation of gauge-theory quench dynamics. (**A**) Schematic nonequilibrium evolution to the steady state. Under the constrained (gauge) condition, we found that different initial states with the same energy density evolved toward a common thermal state of the gauge theory. (**B**) Quantum simulator for the gauge theory. Matter and gauge fields are represented by occupations of bosonic atoms in an optical superlattice. Charges are illustrated as red (positive) and blue (negative) circles, and electric flux is shown as yellow arrows. On matter sites, the presence of an atom signals a corresponding charge in the gauge theory. To illustrate Gauss's law, we indicated locally gauge-invariant configurations around even (green boxes) and odd (blue boxes) matter sites (see also fig. S1). (**C**) Evolution of the matter density measured by in situ imaging. Top: Starting from the initial state with unity-filled matter sites $\langle \hat{n}_{\text{matter}} \rangle = 1$ (see inset), we observed a fast decay of the matter density $\langle \hat{n}_{\text{matter}} \rangle$ for "violent" quenches (m/κ = 0) in our 71-site quantum simulator. Bottom: Evolution of matter density (averaged over 36 matter sites of the superlattice). Error bars indicate SDs.

[1]Hefei National Research Center for Physical Sciences at the Microscale and School of Physical Sciences, University of Science and Technology of China, Hefei 230026, China. [2]School of Physics, University of Science and Technology of China, Hefei 230026, China. [3]Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, 69120 Heidelberg, Germany. [4]CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, China. [5]INO-CNR BEC Center and Department of Physics, University of Trento, Via Sommarive 14, I-38123 Trento, Italy. [6]Institute for Theoretical Physics, Ruprecht-Karls-Universität Heidelberg, 69120 Heidelberg, Germany. [7]Institut für Experimentalphysik, Universität Innsbruck, A-6020 Innsbruck, Austria. [8]Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China.
*Corresponding author. Email: zhouzhy@mail.ustc.edu.cn
†These authors contributed equally to this work.
‡Present address: Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China.

tunneling strength J, staggered potential parameter δ, linear potential Δ, and on-site interaction U, as indicated in Fig. 1B. We used a Jordan-Wigner transformation to replace the fermionic fields in Eq. 1 with bosonic atoms [see (*23*) for derivational details].

We kept matter and gauge fields as dynamical degrees of freedom each represented by appropriate site occupations of atoms in an optical superlattice. Gauge symmetry is enforced by suitable energy penalties constraining the system to a gauge-invariant subspace of the quantum simulator (*24*–*26*). For $J \ll \delta, U$, and a linear potential $\Delta = 57$ Hz, we suppressed both direct and long-range tunneling and realized the gauge theory at second-order in perturbation theory (*22*). We identified the gauge-invariant interaction with a correlated annihilation of two atoms on neighboring matter sites to form a doublon on the gauge link in between (and reverse) (Fig. 1B). The mass of the fermion pair is set by the energy balance of this process as $2m = 2\delta - U$ and the interaction strength is given by $\kappa \approx 8\sqrt{2}J^2/U$ close to resonance ($m \sim 0$).

To describe the nonequilibrium evolution of a gauge theory, it is essential to also respect the gauge symmetry in the initial state. In Fig. 1, we show examples of such initial states, which can be prepared in the present apparatus (*8*). We started the experiment with an array of *36* near-unity-filling chains of ^{87}Rb atoms in the hyperfine state $F = 1$, $m_F = -1$. The individual chains extend over 71 sites of an optical superlattice, which is formed by the superposition of a short lattice (spacing $a_s = 383.5$ nm) and a long lattice (spacing $a_l = 767$ nm). Using the full tunability of superlattice configurations and the recently developed spin-dependent addressing technique (*27*), we removed all atoms on odd (gauge) sites, rendering only the even (matter) sites singly occupied in the initial state. The resulting state corresponds to the ground state of Eq. 1 for $\kappa = 0$ and $m < 0$, and is characterized by empty gauge sites and unity filling on the matter sites $\langle \hat{n}_{\text{matter}} \rangle = 1$, where $\langle \hat{n}_{\text{matter}} \rangle = \sum_{j \in m} \langle \hat{n}_j \rangle / L_{\text{m}}$ is the average number of bosonic atoms over the L_{m} even sites.

After the initial-state preparation, the atoms were isolated in deep lattice wells ($J, \kappa \approx 0$). To initiate the dynamics, we first tuned the superlattice configuration such that potential minima of the two lattices were aligned, creating the staggered potential. The quench was then initiated by tuning the laser intensities to realize the desired values of κ and m, which can be chosen from a broad range. Subsequently, the system undergoes coherent many-body oscillations. After a certain evolution time, we rapidly ramped up the lattice depth along the x-axis to $60E_r$ within 0.1 ms to freeze the dynamics, where $E_r = h^2/\left(8m_{\text{Rb}}a_s^2\right)$ is the recoil energy, m_{Rb} the atom mass, and h Planck's constant. We then used the same site-selective addressing technique and read out $\langle \hat{n}_{\text{matter}} \rangle$ with in situ absorption imaging. Each data point was measured by averaging more than six realizations of the experiment. We show corresponding in situ experimental data in Fig. 1C for evolution times $t \leq 150$ ms, with $\kappa = 14.5$ Hz and $m = 0$. For a broad range of model parameters, we observed that the system relaxed toward a steady state after only a few oscillations. The oscillation frequency was mildly affected by the inhomogeneous Gaussian profile of the optical trap toward the edges ($\Delta U \sim 10$ Hz). Overall, the system retained a high degree of homogeneity throughout the tractable evolution times, as demonstrated in Fig. 1C.

In Fig. 2, we show the system evolution for the most "violent" quench to $m/\kappa = 0$, corresponding to $U \approx 2\delta$. In Fig. 2A, the real-time dynamics at various microscopic Bose-Hubbard parameters, which all map to the same m/κ but with different strengths of the gauge constraint (cases 1 to 3), is compared with theoretical estimates. The top panel shows a result for the ideal gauge field dynamics obtained through exact diagonalization of the Hamiltonian in Eq. 1 for a smaller system with 18 matter sites. The bottom panels give the experimental results for the observable along with numerical estimates based on the time-dependent density matrix renormalization group (*t*-DMRG) (*28*, *29*) for a Bose-Hubbard chain of 32 sites, which show good agreement. In the gauge-theory regime (case 1), we used the damped-sine fitting at later times (*23*) to extract the damping rate γ, which was found to be $\gamma^{-1} = 63 \pm 9$ ms (experiment) and 64.4 ± 0.4 ms (*t*-DMRG). Earliest times can be sensitive to small differences in initial conditions.

The different levels of constrained dynamics were realized by tuning the Bose-Hubbard parameters from $\delta/J = 1$ (case 3) to $\delta/J = 16$ (case 1). This is reflected in the gauge violation η, which tends to zero in the gauge-theory regime (case 1). It is defined as the odd atom number expectation value on gauge sites, $\eta = \sum_{j \in g} \langle \hat{n}_j \bmod 2 \rangle / L_g$, where L_{g} is the number of odd (gauge) sites (*23*). We measured this probability by removing pairs of atoms in the same well with a photoassociation laser, followed by selectively addressing the gauge links for imaging. We used η as a measure to validate our quantum simulation of the gauge theory, finding a controlled decrease from large violations in case 3 toward $\eta \approx 0$ in case 1 (Fig. 2, A to D). To characterize the dynamics as a function of the staggering parameter δ/J, which is used to enforce the gauge constraint, we extracted the oscillation frequency of the matter density with a damped-sine fit. In Fig. 2B, our results show a fast approach toward the gauge theory upon increasing δ/J.

We further investigated the role of the gauge constraint in the relaxation dynamics of the gauge theory by considering quenches to nonzero values of the mass m. This amounts to regions away from the resonance line characterized by $U = 2\delta$ (Fig. 2C), where the annihilation of fermion pairs is strongest. For $m = -0.8\kappa$, the resulting time evolution is displayed in Fig. 3 with both weakly ($\delta/J \sim 2$) and strongly constrained ($\delta/J \sim 11$) dynamics shown. Here and in the following, we focus on a region of interest of 50 chains each with an extent of 50 sites. This mitigates the effects of a slightly inhomogeneous trap. We compared the strongly constrained dynamics with the thermal prediction of the gauge theory, finding agreement within experimental precision at late times. By contrast, the unconstrained system evolves toward a very different state, characterized by a thermal ensemble of the Bose-Hubbard system away from the gauge-theory regime. In Fig. 3, the thermal predictions have been obtained from a numerical evaluation of the corresponding microcanonical and canonical ensembles. We extracted the temperatures from the latter, as shown in Fig. 3, by fixing their (conserved) energy density to that of the pure initial state (*23*).

As a next step, we investigated the role of the initial state in the thermalization dynamics of the gauge theory. If the system approaches thermal equilibrium, then the late-time behavior is entirely characterized by conserved quantities. The tunability of the Hubbard parameters allowed us to access a broad range of gauge-symmetric initial states with an adiabatic ramp, ranging from the fully filled state ($\langle \hat{n}_{\text{matter}} \rangle = 1$) to states in which a large fraction of fermion pairs have been annihilated ($\langle \hat{n}_{\text{matter}} \rangle \approx 0.21$) (*8*) (Fig. 4A). For the quench dynamics, we compared initial states with the same energy density with respect to the quench Hamiltonian (Fig. 4B). To achieve this, we numerically followed the experimental sequence and determined suitable ramp times as outlined in Fig. 4A for two values of $m/\kappa = -0.8$ and $m/\kappa = 0$ (*23*). For $m/\kappa = 0$, the initial states with the same energy density were prepared with the ramp times 1.2 and 20 ms, corresponding to $\langle \hat{n}_{\text{matter}} \rangle = 1$ and $\langle \hat{n}_{\text{matter}} \rangle \approx 0.21$, respectively. For $m/\kappa = -0.8$, the ramp times were 0 and 6.8 ms, corresponding to the initial states of $\langle \hat{n}_{\text{matter}} \rangle = 1$ and $\langle \hat{n}_{\text{matter}} \rangle \approx 0.71$. The ensuing dynamics were characterized by transient many-body oscillations, in which the different initial states relax to a common steady state at long times (Fig. 4, C and D). During the evolution, the information about initially different matter densities is seen to be effectively lost in the quantum many-body system. We again found the long-time steady states to be well described by gauge-invariant thermal ensembles with the same conserved

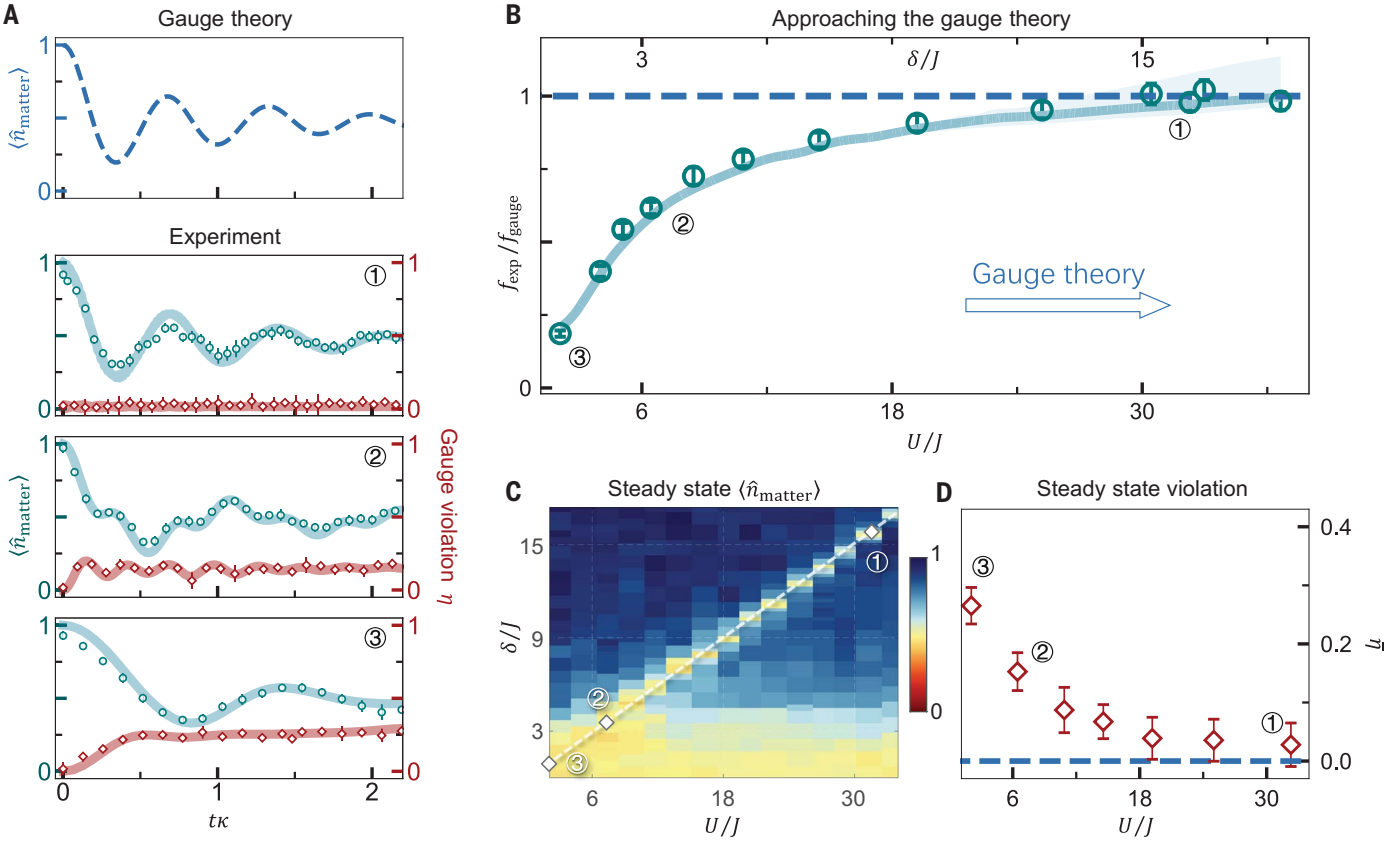


Fig. 2. Controlled approach to gauge-theory dynamics. (**A**) Time evolution in the quantum simulator with and without gauge-theory constraint. The experiment describes the ideal gauge-theory dynamics (top panel) well for case 1 ($U, \delta \gg J$), whereas cases 2 and 3 deviate from these dynamics. The gauge violation, defined as a forbidden odd occupation of gauge sites, decreased from case 3 to case 1 and remained bounded throughout the tractable evolution dynamics. Solid curves are numerical t-DMRG results (*23*). (**B**) Frequency of many-body oscillations. The experiment (f_{exp}) showed a fast approach toward the gauge theory (f_{gauge}) for increasing staggering strength δ. The oscillation frequency was extracted as the lower frequency of a fitted dual-frequency damped-sine function [see (*23*) for details]. (**C**) Two-dimensional parameter space. The late-time (t = 120 ms) matter density was independently scanned over two Hubbard parameters. The dashed white line indicates the resonance condition ($m = 0 \leftrightarrow U = 2\delta$) applied to all the quench evolution data in the figure. (**D**) Steady-state gauge violation. The time-averaged gauge violation $\bar{\eta}$ falls off toward zero for the strongly constrained system with $U/J \gg 1$.

charges as the initial state. We observed these thermalization dynamics for different values of couplings in the gauge theory. Specifically, for m = 0, our initial states were distributed symmetrically around the center of the energy spectrum (*23*). In this case, there was a fast effective loss of initial-state memory, and the experiment relaxed to the steady-state value of the infinite-temperature state (Fig. 4, B and C).

Effectively irreversible behavior, such as thermalization from an underlying reversible time evolution, emerges in general in non-integrable models for local observables and typical initial states (*4*). Despite the non-integrability of the U(1) quantum link model (*14*, *30*, *31*), certain fine-tuned quenches give rise to weak ergodicity breaking because of the presence of special eigenstates in the spectrum of the quench Hamiltonian (*14*, *30*). This could manifest in persistent oscillations around the thermal-ensemble prediction (*13*). Currently, the level of control in our experiment limits us from probing such possible behavior, in part because of an inherent residual inhomogeneity across the lattice. Thus, in our experiment, we observed equilibration to close to the thermal equilibrium value for all of the quenches that we performed.

Our work creates a pathway for addressing emergent dynamical phenomena in gauge theories, such as the Schwinger effect (*32*), dynamical topological quantum phase transitions (*33*), and string breaking (*34*, *35*) in strong fields (*36*). The approach lays also the foundations for the exploration of more complex, higher-dimensional gauge theories using state-of-the-art quantum technology (*37*). An important next step toward applications for gauge theories such as quantum electrodynamics, or maybe even quantum chromodynamics, is a faithful extension of the discrete quantum-link representation toward continuous variables (*9*, *38*, *39*). To this end, current

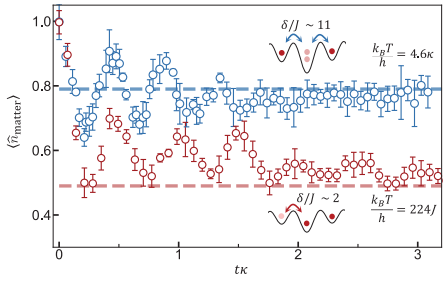


Fig. 3. Thermalization dynamics with and without the gauge-symmetry constraint. We investigated the real-time evolution of two datasets toward a late-time steady state for constrained (blue) and unconstrained (red) dynamics. Dashed lines show exact-diagonalization predictions from canonical thermal ensembles for the gauge theory (blue) and the Bose-Hubbard model (red) with the same energy density (*23*). Insets show relevant processes with (top) and without (bottom) the gauge-symmetry constraint.

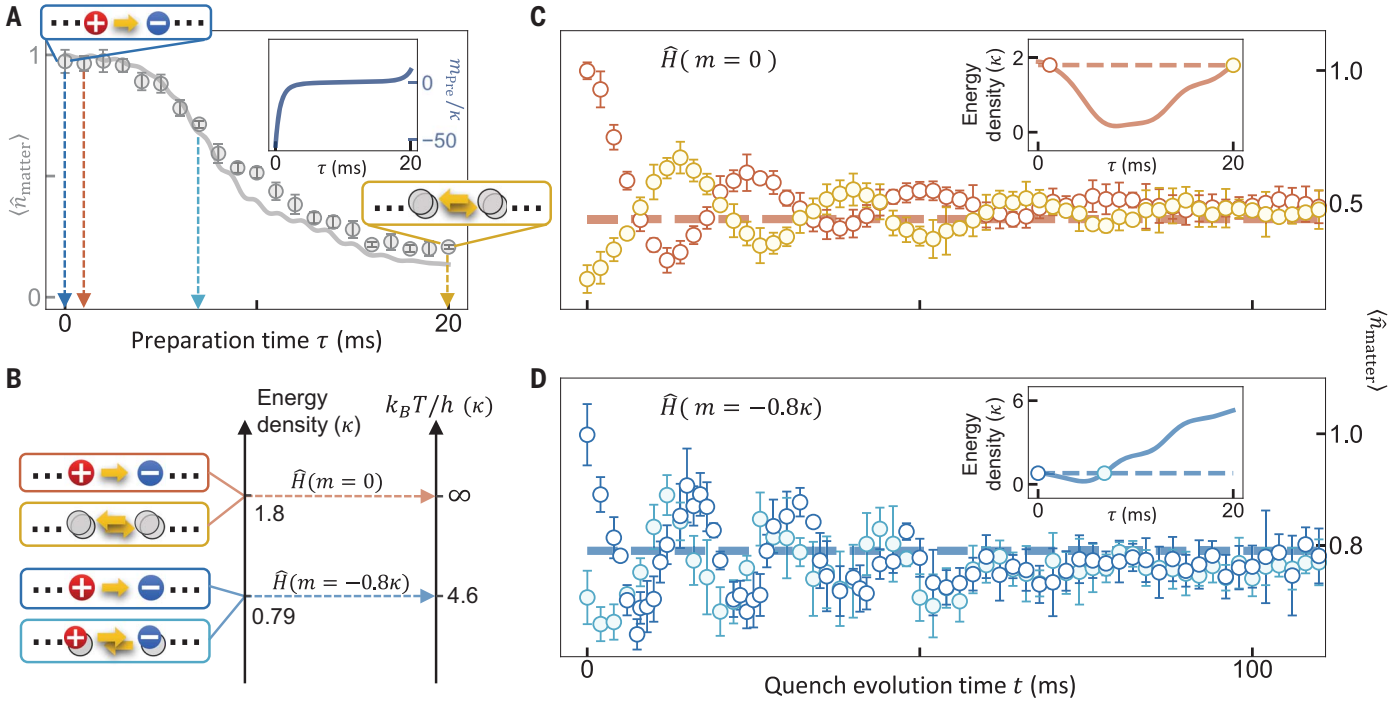


Fig. 4. Effective loss of initial-state information. (**A**) State preparation. Evolution of the matter density from the "fully matter-filled" state ($\langle\hat{n}_{\text{matter}}\rangle = 1$, blue box, left) to almost "matter-empty" state ($\langle\hat{n}_{\text{matter}}\rangle \approx 0.21$, yellow box, right) for the adiabatic ramp with preparation time τ and corresponding mass parameter m_{Pre}/κ as shown in the inset. (**B**) Schematic of the evolution toward thermal equilibrium. For each of two sets of quench parameters ($m = 0$ and $m = -0.8\kappa$), we chose two initial states with equal energy density. The resulting steady states in the wake of the quenches starting in these two initial states were then compared with a canonical thermal ensemble with the temperature determined from the energy density (*23*). Here, all energy densities are plotted with respect to the ground state of the evolution Hamiltonian. (**C** and **D**) Relaxation. Shown are the thermalization dynamics for the chosen quench parameters and initial states [shown in (B)]. Experimental data are compared with predictions from corresponding gauge theory thermal ensembles (dashed lines) at temperatures $k_BT = \infty$ (top) and $k_BT = 4.6\kappa$ (bottom). The insets show the energy density evolution during state preparation, and the circles mark the chosen initial states.

implementation schemes should be extended to higher spin representations and scalable, higher-dimensional setups (*40*).

REFERENCES AND NOTES

1. M. C. Bañuls *et al.*, *Eur. Phys. J. D* **74**, 165 (2020).
2. V. Kasper, F. Hebenstreit, J. Berges, *Phys. Rev. D* **90**, 025016 (2014).
3. J. Berges, J. Cox, *Phys. Lett. B* **517**, 369–374 (2001).
4. M. Rigol, V. Dunjko, M. Olshanii, *Nature* **452**, 854–858 (2008).
5. J. Berges, M. P. Heller, A. Mazeliauskas, R. Venugopalan, *Rev. Mod. Phys.* **93**, 035003 (2021).
6. E. A. Martinez *et al.*, *Nature* **534**, 516–519 (2016).
7. C. Kokail *et al.*, *Nature* **569**, 355–360 (2019).
8. B. Yang *et al.*, *Nature* **587**, 392–396 (2020).
9. A. Mil *et al.*, *Science* **367**, 1128–1130 (2020).
10. C. Schweizer *et al.*, *Nat. Phys.* **15**, 1168–1173 (2019).
11. F. Görg *et al.*, *Nat. Phys.* **15**, 1161–1167 (2019).
12. H.-N. Dai *et al.*, *Nat. Phys.* **13**, 1195–1200 (2017).
13. H. Bernien *et al.*, *Nature* **551**, 579–584 (2017).
14. F. M. Surace *et al.*, *Phys. Rev. X* **10**, 021041 (2020).
15. N. Klco *et al.*, *Phys. Rev. A* **98**, 032331 (2018).
16. W. A. de Jong *et al.*, arXiv:2106.08394 [quant-ph] (2021).
17. S. Chandrasekharan, U. J. Wiese, *Nucl. Phys. B* **492**, 455–471 (1997).
18. M. Levin, X. G. Wen, *Rev. Mod. Phys.* **77**, 871–879 (2005).
19. M. Hermele, M. P. A. Fisher, L. Balents, *Phys. Rev. B* **69**, 064404 (2004).
20. J. B. Kogut, *Rev. Mod. Phys.* **55**, 775–836 (1983).
21. R. Brower, S. Chandrasekharan, U.-J. Wiese, *Phys. Rev. D* **60**, 094502 (1999).
22. J. Kogut, L. Susskind, *Phys. Rev. D* **11**, 395–408 (1975).
23. See the supplementary materials.
24. E. Zohar, J. I. Cirac, B. Reznik, *Phys. Rev. Lett.* **110**, 055302 (2013).
25. J. C. Halimeh, P. Hauke, *Phys. Rev. Lett.* **125**, 030503 (2020).
26. J. C. Halimeh, H. Lang, J. Mildenberger, Z. Jiang, P. Hauke, *PRX Quantum* **2**, 040311 (2021).
27. B. Yang, H.-N. Dai, A. Reingruber, Z.-S. Yuan, J.-W. Pan, *Phys. Rev. A* **96**, 011602 (2017).
28. U. Schollwöck, *Ann. Phys.* **326**, 96–192 (2011).
29. I. P. McCulloch, "The Matrix product toolkit" (2021); https://people.smp.uq.edu.au/IanMcCulloch/mptoolkit/.
30. C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, Z. Papić, *Nat. Phys.* **14**, 745–749 (2018).
31. J. C. Halimeh, M. V. Damme, T. V. Zache, D. Banerjee, P. Hauke, arXiv:2112.04501 [cond-mat.quant-gas] (2021).
32. T. V. Zache *et al.*, *Quantum Sci. Technol.* **3**, 034010 (2018).
33. Y.-P. Huang, D. Banerjee, M. Heyl, *Phys. Rev. Lett.* **122**, 250401 (2019).
34. S. Banerjee, *Phys. Rev. Lett.* **109**, 010402 (2012).
35. F. Hebenstreit, J. Berges, D. Gelfand, *Phys. Rev. Lett.* **111**, 201601 (2013).
36. N. B. Narozhny, A. M. Fedotov, *Eur. Phys. J. Spec. Top.* **223**, 1083–1092 (2014).
37. E. Zohar, *Philos. Trans. A Math. Phys. Eng. Sci.* **380**, 20210069 (2022).
38. M. Prüfer *et al.*, *Nat. Phys.* **16**, 1012–1016 (2020).
39. T. V. Zache, T. Schweigler, S. Erne, J. Schmiedmayer, J. Berges, *Phys. Rev. X* **10**, 011020 (2020).
40. R. Ott, T. V. Zache, F. Jendrzejewski, J. Berges, *Phys. Rev. Lett.* **127**, 130504 (2021).
41. Data for: Z.-Y. Zhou, G.-X. Su, J. C. Halimeh, R. Ott, H. Sun, P. Hauke, B. Yang, Z.-S. Yuan, J. Berges, J.-W. Pan, Thermalization dynamics of a gauge theory on a quantum simulator, Harvard Dataverse, Version 1 (2022); https://doi.org/10.7910/DVN/XWUHTS.

ACKNOWLEDGMENTS

We thank D. Banerjee, A. Sen, J.-Y. Desaules, M. Gärttner, A. Hudomal, F. Jendrzejewski, A. Lazarides, M. Oberthaler, Z. Papić, J. Schmiedmayer, C. J. Turner, and T. V. Zache for discussions.

Funding: This work is part of and supported by the National Natural Science Foundation of China (NNSFC grant 12125409); the Innovation Program for Quantum Science and Technology (grant 2021ZD0302000); the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation project ID 273811115-SFB 1225); Germany's Excellence Strategy EXC 2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster); the Anhui Initiative in Quantum Information Technologies; the Chinese Academy of Sciences; Provincia Autonoma di Trento; ERC Starting Grant StrEnQTh (project ID 804305); Google Research Scholar Award ProGauge; and Q@TN (Quantum Science and Technology in Trento). **Author contributions:** Z.-Y.Z., G.-X.S., and H.S. performed the experiments and analyzed the data. J.C.H. and R.O. performed the numerical calculations and comparisons with experimental data. H.S., B.Y., Z.-S.Y., and J.-W.P. developed relevant experimental techniques. P.H., B.Y., and J.B. initiated early discussions on this topic. J.B., P.H., J.C.H., and R.O. developed the theory. Z.-S.Y. and J.-W.P. conceived and supervised the experimental research. All authors contributed to the writing of the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All experimental data and code are available in the Harvard Dataverse database (*41*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl6277
Supplementary Text
Figs. S1 to S5
References (*42*, *43*)

Submitted 28 July 2021; accepted 13 June 2022
10.1126/science.abl6277

WATER STRUCTURE

Visualizing Eigen/Zundel cations and their interconversion in monolayer water on metal surfaces

Ye Tian[1]†, Jiani Hong[1]†, Duanyun Cao[2,3]†, Sifan You[4]†, Yizhi Song[1], Bowei Cheng[1], Zhichang Wang[1], Dong Guan[1], Xinmeng Liu[1], Zhengpu Zhao[1], Xin-Zheng Li[5,6,7,8], Li-Mei Xu[1,6,8], Jing Guo[9]*, Ji Chen[5,8]*, En-Ge Wang[1,6,8,10]*, Ying Jiang[1,6,8,11]*

The nature of hydrated proton on solid surfaces is of vital importance in electrochemistry, proton channels, and hydrogen fuel cells but remains unclear because of the lack of atomic-scale characterization. We directly visualized Eigen- and Zundel-type hydrated protons within the hydrogen bonding water network on Au(111) and Pt(111) surfaces, using cryogenic qPlus-based atomic force microscopy under ultrahigh vacuum. We found that the Eigen cations self-assembled into monolayer structures with local order, and the Zundel cations formed long-range ordered structures stabilized by nuclear quantum effects. Two Eigen cations could combine into one Zundel cation accompanied with a simultaneous proton transfer to the surface. Moreover, we revealed that the Zundel configuration was preferred over the Eigen on Pt(111), and such a preference was absent on Au(111).

Hydrated protons are ubiquitous in solutions and involved in a variety of physical, chemical, biological, and energy-related processes (*1*–*9*). The solvation of hydrated proton lies at the heart of acid-base reactions (*10*), enzymatic functions (*11*), proton channels (*12*, *13*), and hydrogen-fuel cells (*7*, *14*). Despite tremendous theoretical and experimental efforts, the nature of hydrated proton in water has been under long-standing and lively debate, mainly because of the lack of comprehensive atomic-level understanding. In particular, the solvation and dynamics of hydrated protons at water-solid interfaces are highly relevant to the key features of electrochemical reactions, such as hydrogen evolution reaction (HER) (*15*, *16*). Whether the hydrated protons are stable and in what forms they exist still remain unclear, hindering a deeper understanding on the detailed reaction pathways on different electrodes (*16*, *17*).

Many different forms of hydrated protons in water have been reported, among which the Zundel cation ($H_5O_2^+$) (*18*) and the Eigen cation ($H_9O_4^+$) (*19*) are the most representative ones. However, the configuration of Zundel and Eigen cations in bulk and interfacial water has been experimentally elusive. Vibrational spectroscopy, which is sensitive to hydrogen (H) bonding strength and dynamics, has been used to identify the molecular structure of hydrated proton (*20*–*24*). Nevertheless, the spectral approach is inefficient because of the transient nature of hydrated protons and extremely diffuse vibrational signatures. Although this problem can be largely avoided for probing protonated gas-phase clusters (*25*–*27*), the obtained knowledge of an isolated cluster is not necessarily applicable to extended H bonding networks and interfacial systems, in which multiple interactions usually compete.

Recently, qPlus-based atomic force microscopy (AFM) (*28*) with a carbon monoxide (CO)–terminated tip has been successfully applied to probe interfacial water and hydrated alkali ions with submolecular resolution in a nearly noninvasive manner (*29*–*31*), on the basis of delicate high-order electrostatic forces. The proton affinity of individual surface hydroxyl was also directly quantified with an OH-terminated tip (*32*). However, imaging and identifying hydrated protons within the H bonding network of water remain a great challenge because of the high similarity between the hydronium (H_3O^+) and the water molecule (H_2O). In this work, we further improved the resolution and sensitivity of the qPlus-AFM [figs. S1 and S2 and supplementary materials (SM)], so that the Zundel and Eigen cations could be directly visualized and distinguished within the monolayer water on metal surfaces at 5 K under ultrahigh vacuum.

Scanning tunneling microscopy (STM) images of the coadsorbed deuterium (D) atoms and D_2O molecules on Au(111) surface are shown in Fig. 1A (SM, materials and methods). The substitution between H and D had a negligible effect on the structures of hydronium-water layers. Most of the results shown in the current work were done with D because of the higher stability of D-doped D_2O layer under the scanning. The D atoms were ionized by transferring electron charge to the Au substrate and water molecules (figs. S3 and S4). The ionized D^+ and D_2O molecules could self-assemble into a two-dimensional (2D) island with a monolayer height (Fig. 1A, bottom right, and fig. S5), which exhibited a honeycomb structure commensurate with the Au(111) lattice (Fig. 1A, bottom left). However, it is challenging to identify hydronium ions (D_3O^+) solvated in the H bonding network of water from the STM images. The height-dependent AFM images and simulations of the 2D hydronium-water structure are shown in Fig. 1, B to D. At the large tip height, the hydronium ions were visualized as individual protrusions with random distribution mainly through the high-order electrostatic force (Fig. 1B, top, and fig. S6) (*29*, *31*). The zoomed-in AFM image clearly indicates that each bright protrusion was surrounded by three symmetric depressions (Fig. 1B, orange dashed curves). Such a feature is consistent with the structure of D_3O^+, in which the negatively charged O and positively charged D yield opposite force contrast. When the tip approached the surface, so that the Pauli repulsive force dominated (*33*, *34*), the bright protrusion evolved into a three-pointed star (Fig. 1C, top). At the smallest tip height, the hexagonal H bonding skeleton of water appeared, and the hydronium ion was imaged as a depression feature owing to the relaxation of the CO tip (Fig. 1D, top).

Although the distribution of hydronium ions has no long-range order, the distance between them was typically 505 and 569 pm, corresponding to the meta-site and the para-site of the hexagonal water ring, respectively (fig. S7). To gain insights into the structure of the solvated hydronium, we constructed a periodic structure that consists of para-hydronium (Fig. 1E). Such a structure was energetically stable (fig. S8) and could be considered as the self-assembly of Eigen cations ($D_9O_4^+$), in which one flat hydronium ion at the center is H bonded with three water molecules (*19*).

[1]International Center for Quantum Materials, School of Physics, Peking University, Beijing 100871, China. [2]Beijing Key Laboratory of Environmental Science and Engineering, School of Materials Science and Engineering, Beijing Institute of Technology, Beijing 100081, China. [3]Beijing Institute of Technology Chongqing Innovation Center, Chongqing 401120, China. [4]Institute of Functional Nano and Soft Materials, Jiangsu Key Laboratory for Carbon-Based Functional Materials and Devices, Soochow University, Suzhou 215123, China. [5]School of Physics, Peking University, Beijing 100871, China. [6]Collaborative Innovation Center of Quantum Matter, Beijing 100871, China. [7]State Key Laboratory for Mesoscopic Physics and Frontiers Science Center for Nano-optoelectronics, School of Physics, Peking University, Beijing 100871, China. [8]Interdisciplinary Institute of Light-Element Quantum Materials and Research Center for Light-Element Advanced Materials, Peking University, Beijing 100871, China. [9]College of Chemistry, Beijing Normal University, Beijing 100875, China. [10]Songshan Lake Materials Lab, Institute of Physics, CAS and School of Physics, Liaoning University, Shenyang 110036, China. [11]CAS Center for Excellence in Topological Quantum Computation, University of Chinese Academy of Sciences, Beijing 100190, China.
***Corresponding author. Email: jguo1294@bnu.edu.cn (J.G.); ji.chen@pku.edu.cn (J.C.); egwang@pku.edu.cn (E.-G.W.); yjiang@pku.edu.cn (Y.J.)**
†These authors contributed equally to this work.

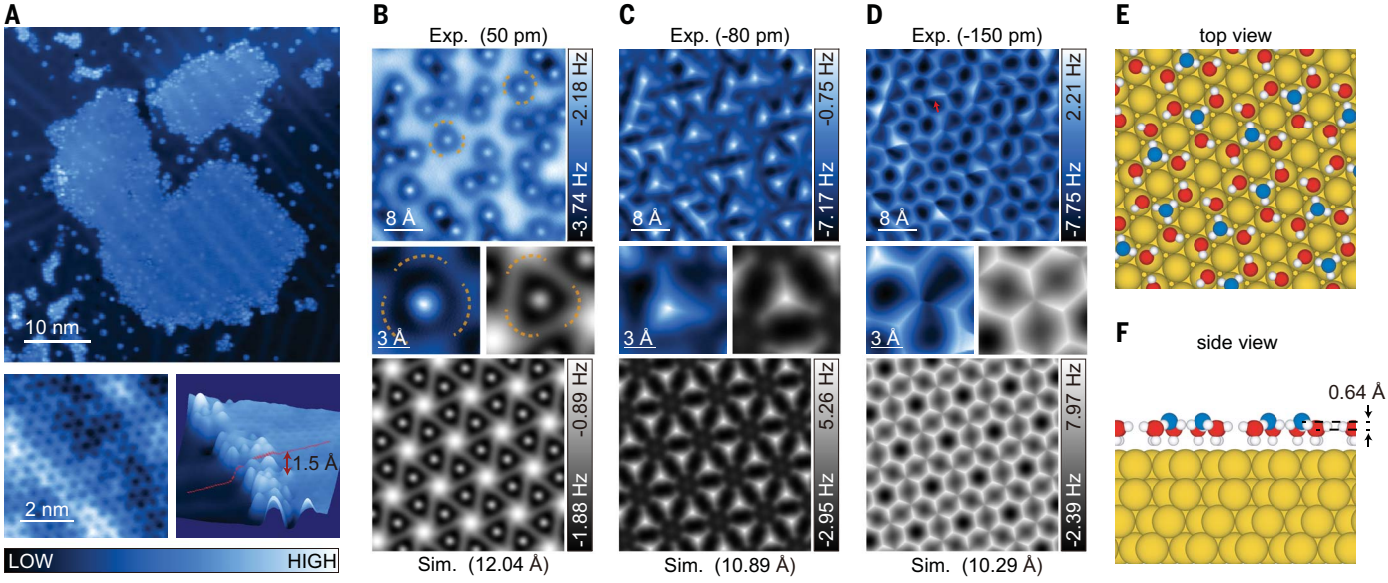


Fig. 1. Atomic structure of self-assembled Eigen-water monolayer on Au(111) surface. (**A**) Overview STM image of the Eigen-water layer formed on Au(111) surface after coadsorption of D atoms and water molecules (D_2O) and subsequently annealing at 120 K. (Bottom) Zoom-in STM images of the hydronium-water island, showing the hexagonal honeycomb structure (left) with the height of ~1.5 Å (right). STM set point, 100 mV and 50 pA. (**B** to **D**) Height-dependent AFM images and simulations of the hydronium-water layer, obtained at the tip heights of (B) (top) 50 pm and (bottom) 12.04 Å, (C) (top) −80 pm and (bottom) 10.89 Å, and (D) (top) −150 pm and (bottom) 10.29 Å. (Middle) Zoom-in AFM images and the simulations of a single Eigen cation in the 2D hydronium-water layer. The yellow dashed curves indicate three symmetric depressions induced by the positively charged hydronium. (**E**) Top and (**F**) side views of the schematic structures of the periodic Eigen-water monolayer with Eigen cations residing at the para-site of the hexagonal water ring. The height difference between hydronium ion and water molecule is indicated in (F). The tip heights in AFM images are referenced to the STM set point on the Eigen-water layer (100 mV and 50 pA). The tip heights in AFM simulations are defined as the vertical distance between the apex atom of the metal tip and the outmost atom of Au substrate. Au, H, and O atoms in the H-down water molecules are indicated with yellow, white, and red spheres, respectively. O atoms in the hydronium are indicated with blue spheres.

Density functional theory (DFT) calculation revealed that the central hydronium ion was ~64 pm higher than the neighboring water molecules, which were in an H-down configuration because of the negatively charged surface (Fig. 1F). The simulated AFM images (Fig. 1, B to D, bottom, and fig. S6) agreed well with the experimental ones (Fig. 1, B to D, top, and fig. S6), further validating the assignment of the Eigen cations. The introduction of meta-hydronium may break the long-range order and led to local defects (Fig. 1D, top, red arrow).

A more ordered structure could form when dosing D at higher coverages (Fig. 2). The STM image (Fig. 2A) showed a similar honeycomb structure as the one in Fig. 1A, but with a larger apparent height (Fig. 2A, bottom right, and fig. S5) and a smaller lattice constant (table S1). The AFM images (Fig. 2, B to D, and fig. S6) at the large and intermediate tip heights exhibit a dimer-like structure with a long-range 3 × 3 periodicity (Fig. 2, B and C, top), with a domain size up to 900 nm^2 (fig. S10). At the smallest tip height, a honeycomb structure was observed (Fig. 2D, top). It is straightforward to infer that such a structure may arise from the dimerization of the ortho-site hydronium ions, so that the two flat water molecules share an extra D^+ in a symmetric H bonding configuration (Fig. 2E), corresponding to the Zundel cation ($D_5O_2^+$) (*18*). The Zundel cation together with four neighboring H-down water molecules formed a basic unit ($D_{13}O_6^+$), which then self-assembled into a perfect 3 × 3 structure (Fig. 2E and fig. S9).

To further explore the stability of the Zundel-water monolayer, we performed detailed path integral molecular dynamics (PIMD) simulations based on DFT calculations, which include the nuclear quantum effects (NQEs) explicitly (SM, materials and methods, and table S2). These results showed that the formation of the Zundel cation was closely related to the NQEs (*1*), which significantly promoted the proton delocalization between the water molecules (Fig. 2E and figs. S11 to S13). The Zundel cation in the configuration based on the PIMD had a ~72 pm larger height than that of the H-down water molecules (Fig. 2F), which is similar to the case of Eigen configuration. The AFM simulation of such a configuration (Fig. 2, B to D, bottom, and fig. S6) nicely reproduced the AFM features observed in the experiment (Fig. 2, B to D, top, and fig. S6). The evidence of the symmetric H bond could be clearly seen in Fig. 2B, in which the depression feature arising from the shared D^+ was located exactly at the center of each dimer (Fig. 2B, middle, red arrows). Without considering the NQEs, the Zundel cation became unstable and relaxed to the Eigen configuration, which showed a distinct asymmetric dimer feature in the AFM simulation (fig. S14). Consequently, the perfect agreement between the experimental results and the simulated AFM images indicated that the resolution of our qPlus-AFM was high enough to distinguish between the Eigen and Zundel cations, in which the position of protons along the H bond only differed by ~0.2 Å.

Although both the Eigen-type ($H_9O_4^+$) and Zundel-type ($H_{13}O_6^+$) clusters are stable species in the gas phase (*26*), their self-assembly at the interface has not been reported before. The lattice constant of the Zundel-water layer is smaller than that of the Eigen-water layer (table S1), suggesting that the inter–Zundel-cluster H bonding strength in the Zundel-water layer should be larger than that in the Eigen-water layer. Such a difference may arise from the stronger intracluster H bonds within the Eigen cluster, which weakened the intercluster H bonds (table S3). We have observed isolated $D_9O_4^+$ clusters on the surface, and isolated $D_{13}O_6^+$ were absent (fig. S15). Furthermore, we

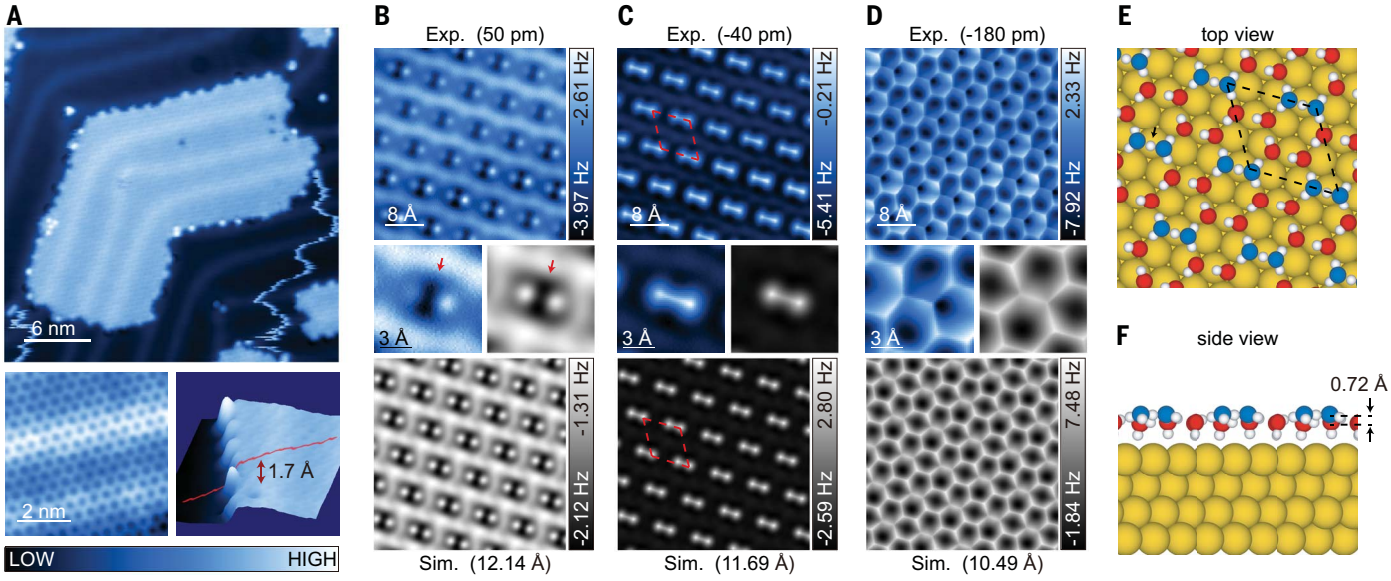


Fig. 2. Atomic structure of self-assembled Zundel-water monolayer on Au(111) surface. (**A**) Overview STM image of the Zundel-water island. (Bottom) Zoom-in STM images, showing the hexagonal honeycomb structure (left) with a height of ~1.7 Å (right). STM set point, 100 mV and 50 pA. (**B** to **D**) Height-dependent AFM images and simulations of the Zundel-water layer, obtained at the tip heights of (B) (top) 50 pm and (bottom) 12.14 Å, (C) (top) −40 pm and (bottom) 11.69 Å, and (D) (top) −180 pm and (bottom) 10.49 Å. (Middle) Zoom-in AFM images and the simulations of a Zundel-type cation in the 2D Zundel-water layer. (**E**) Top and (**F**) side views of the schematic configurations of the Zundel-water layer with a long-range 3 × 3 periodicity obtained with PIMD simulations. The height difference between hydronium ion and water molecule is indicated in (F). The rhombuses in (C) and (E) indicate the 3 × 3 unit cells. The asymmetry of dimers in (B) was caused by the tip asymmetry (fig. S10). The red arrows in (B) indicate the depression feature at the center of each dimer. The black arrow in (E) indicates the shared proton in the Zundel cation. The tip heights in AFM images are referenced to the STM set point on the Zundel-water layer (100 mV and 50 pA). The tip heights in AFM simulations have the same reference and definition as shown in Fig. 1. Au, H, and O atoms in the H-down water molecules are indicated with yellow, white, and red spheres, respectively. O atoms in the Zundel cation are indicated with blue spheres.

found that the interaction between the Zundel-water layer and substrate was weaker than that with the Eigen-water layer (table S3), which was consistent with the observed larger height of the Zundel-water layer (Figs. 1A and 2A and fig. S5). The height difference may result from the weaker inter-Eigen H bonds and the larger density of hydronium in the Eigen-water layer, which could cause a stronger electrostatic attraction between the Eigen-water layer and the substrate (fig. S4 and table S3).

To explore the dynamics of hydronium ions at the interface, which is essential for understanding the HER process (*35*), tip manipulation experiments were carefully performed. The interconversion between the Eigen and Zundel cations induced by the tip is shown in Fig. 3. We applied a voltage pulse on one Zundel cation in the 3 × 3 structure (Fig. 3, A and B, dashed ellipse), leading to the formation of two Eigen cations at the meta-sites (Fig. 3, D and E). The water molecules shared by two hydronium ions after the conversion are indicated with arrows in Fig. 3, E and F. In particular, the water molecule (Fig. 3, E and F, yellow arrow) formed a Bjerrum D-type defect with the neighboring water molecule, whose apparent bond length (300 pm) is larger than that of the intact H bonds (280 pm) (*36*). Those water molecules showed different features from the others, which could be reproduced well with AFM simulation (fig. S16). By applying another voltage pulse, the two Eigen cations could convert back to the Zundel cation.

The interconversion between two Eigen cations and one Zundel cation could only be possible when the D^+ transfer occurred not only between the water molecules but also between the water layer and the surface (Fig. 3, G and H). As voltage pulses were applied, the D-down water molecule shared by the two Eigen cations transferred the downward D^+ to the surface and simultaneously accepted the D^+ transferred from the adjacent hydronium, leading to a flat-lying water molecule and an adsorbed D atom on the surface (D*). Then, the flat water molecule accepted a D^+ from the other adjacent hydronium and formed a symmetric H bond, finally resulting in a stable Zundel+D* configuration (figs. S17 and S18). This conversion from two Eigen cations to the Zundel+D* could be also evidenced in the temperature-dependent experiments (fig. S19). We found that an Eigen cation could neither convert to a Zundel cation without involving interfacial proton transfer nor donate its extra proton to the surface without the help of an additional Eigen cation (fig. S18). Therefore, the formation of Zundel cation and interfacial proton transfer were always coupled together. Such a process (from two Eigens to Zundel+D*) actually corresponds to a previously unknown pathway of Volmer step in HER (*37*), which may greatly facilitate the subsequent Heyrovsky reaction (*38*).

To explore how general those results obtained on Au(111) could be, we performed similar investigations on Pt(111), which is more reactive and hydrophilic than Au(111). It is well known that water forms a $\sqrt{37} \times \sqrt{37}$ monolayer on Pt(111) (*39*, *40*), consisting of pentagons, hexagons, and heptagons (Fig. 4A). When the dosed D coverage was low, individual Zundel cations appeared locally in the $\sqrt{37} \times \sqrt{37}$ phase (Fig. 4D, red arrows). As the D coverage increased, the $\sqrt{37} \times \sqrt{37}$ phase gradually transformed into a $\sqrt{57} \times \sqrt{57}$ phase (Fig. 4B), and the density of Zundel cation also increased, accompanied with the occasional appearance of Eigen cation (Fig. 4E, yellow arrows). The $\sqrt{57} \times \sqrt{57}$ phase finally changed to a honeycomb structure at high D coverages (Fig. 4C), where the Zundel cations formed multiple 3 × 3 domains and the Eigen cations

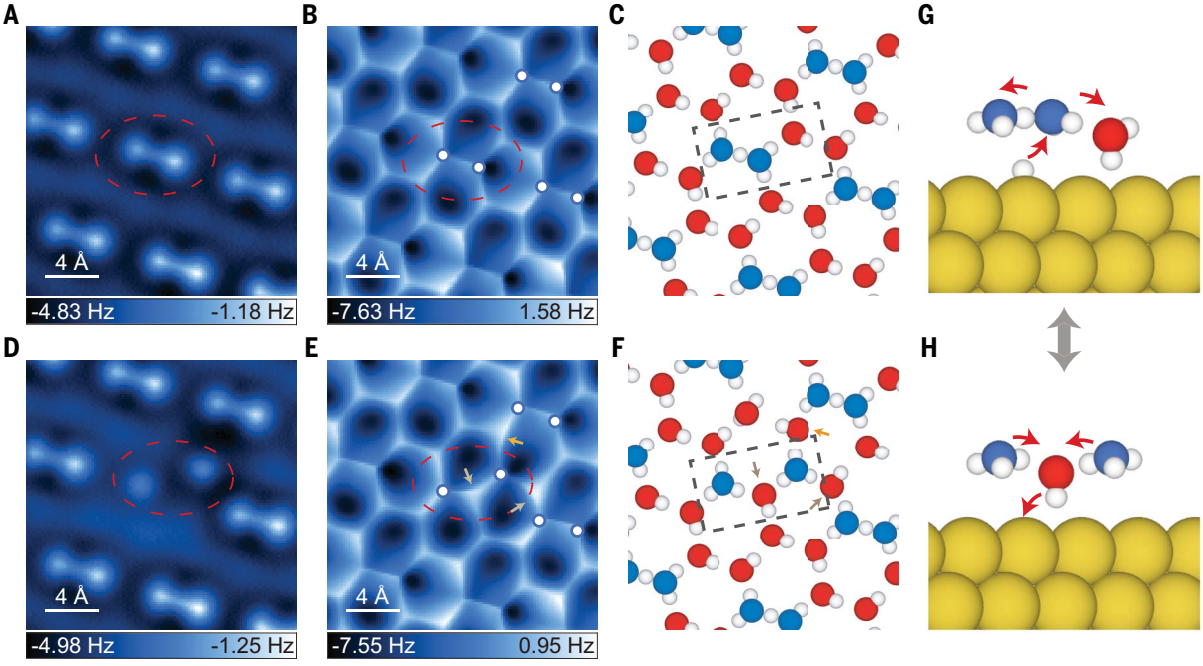


Fig. 3. Tip-induced interconversion between the Eigen and Zundel cations. (**A** to **F**) [(A), (B), (D), and (E)] Height-dependent AFM images and [(C) and (F)] the schematic H-bonding arrangements of [(A) to (C)] the ordered Zundel cations and [(D) to (F)] the tip-induced Eigen cations in the hydronium-water layer. Experimental AFM images were recorded at the tip heights of [(A) and (D)] −50 pm and [(B) and (E)] −180 pm. The red dashed ellipses in (A), (B), (D), and (E) and the black dashed rectangles in (C) and (F) indicate the transition from a Zundel cation (dimer) to two Eigen cations (monomer). The gray and yellow arrows in (E) indicate the AFM features induced by the H-down water molecules shared by the hydroniums at the meta-sites [(F), gray arrows] and the one at the Bjerrum D-type defect [(F), yellow arrow], respectively. (**G** and **H**) Schematic models of the interconversion between the Zundel-type and Eigen-type cations. Au, H, and O atoms in the H-down water molecules are indicated with yellow, white, and red spheres, respectively. O atoms in the Zundel and Eigen cations are indicated with blue spheres. The red arrows indicate the proton transfer pathway during the transition. The tip heights in AFM images are referenced to the STM set point on the Zundel-water layer (100 mV and 50 pA).

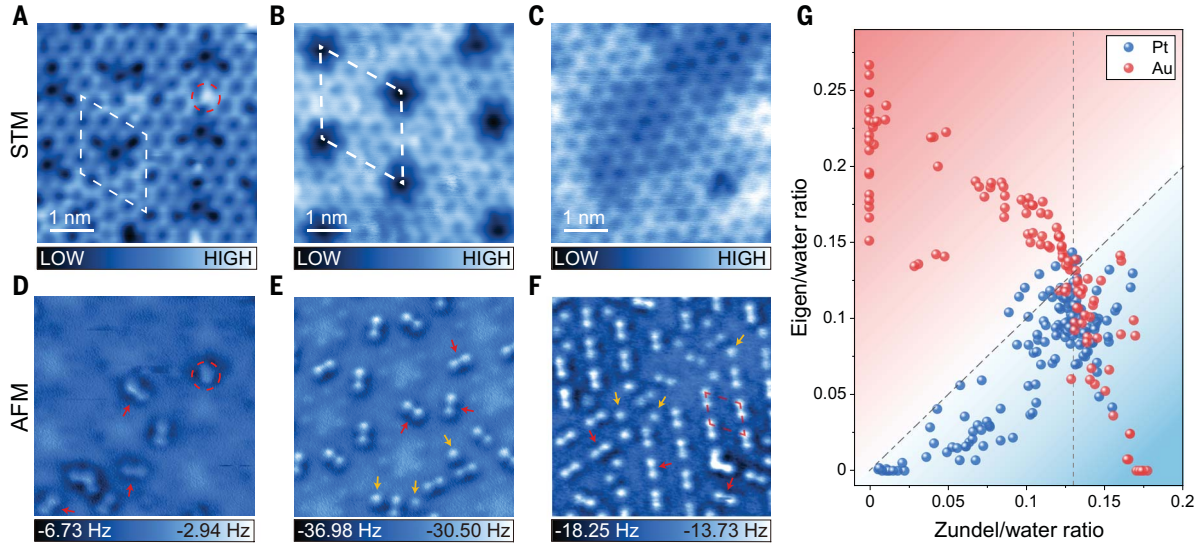


Fig. 4. Hydronium-water overlayer on Pt(111) surface. (**A** to **C**) STM topography and (**D** to **F**) the corresponding AFM images of the hydronium-water overlayer on Pt(111) surface obtained with increasing D coverages. The rhombuses in (A), (B), and (F) indicate the $\sqrt{37} \times \sqrt{37}$, $\sqrt{57} \times \sqrt{57}$, and 3 × 3 unit cells, respectively. STM set points on Pt(111) are (A) 100 mV and 20 pA, (B) 100 mV and 50 pA, and (C) 50 mV and 40 pA. Experimental Δf images were recorded at the tip heights of (D) 80 pm, (E) 100 pm, and (F) 0 pm, respectively. In (D) to (F), the red and yellow arrows indicate the dimer and monomer features, respectively, corresponding to the Zundel- and Eigen-type cations. In (A) and (D), the red dashed circles indicate the D-up water molecule, showing as bright protrusions in both STM and AFM images. (**G**) Density correlation between the Eigen and Zundel on the Au(111) and Pt(111) surfaces at different proton/water ratios (0.01 to 0.45) and different annealing temperature (110 to 145 K), obtained from 35 different samples for Au(111) and 30 different samples for Pt(111). The red and blue shaded areas indicate Eigen- and Zundel-dominant regions, respectively. The vertical dashed line indicates where the Zundel/water ratio equals to 0.13. The diagonal dashed line indicates where the ratios of Zundel/water and Eigen/water are equal.

mainly located at the domain boundaries (Fig. 4F and figs. S20 and S21). We could also achieve the interconversion between two Eigen cations and one Zundel cation on Pt(111) through tip manipulation (fig. S22), suggesting that there was a D* under the Zundel cation (Zundel+D*).

Although the Eigen and Zundel cations could form the similar monolayer structure on Pt(111) and Au(111), there are still some differences between these two surfaces. In Fig. 4G, we plot the density correlation between the Eigen and Zundel cations (figs. S21 and S23), in which two main features could be identified. First, there was a clear transition from the Eigen to Zundel structure on Au(111), and the Zundel was always preferred on Pt(111). The density correlation followed the same behavior for both surfaces when the ratio of Zundel/water was larger than 0.13. Second, the Eigen and Zundel structures could exist even at very low densities on Pt(111), and for Au(111), a minimum density of Eigen and Zundel was needed to stabilize the monolayer. Those distinct features could be understood through the theoretical analysis of thermodynamic stability of the Eigen and Zundel phases in terms of the H chemical potential, which shows that the transition point from the Eigen phase to the Zundel phase on Pt(111) shifts to considerably lower H potential than that on Au(111) (fig. S24). The lattice constant of Pt(111) is smaller than that of the Au(111), thus favoring the delocalization of the extra proton between the water molecules and stabilizing Zundel over Eigen (table S1). The preference of Zundel+D* configuration on Pt(111) could be also related to the stronger Pt-D interaction. In addition, the stronger interaction between the Pt(111) and water could help to stabilize the hydrated protons at small densities.

Our results may provide new insights into the different reaction kinetics of HER on Au(111) and Pt(111) surfaces. On Au(111), the Eigen configuration was dominant, and the Zundel only appeared when the Eigen density became sufficiently large, and the Zundel configuration was always preferred on Pt(111) regardless of the proton density. This suggested that distant hydrated protons were more inclined to combine on Pt(111), thus facilitating the production of H_2. At low Zundel density, the Zundel+H* configuration on Pt(111) may allow efficient H_2 evolution through the Heyrovsky reaction pathway. However, at high Zundel density, the coverage of H* on the Pt(111) was also increased, which promoted the Tafel reaction pathway (*41*, *42*). This feature may provide microscopic insights into the behavior of reaction kinetics on Pt(111) as changing the electrochemical potential (the proton density near the surface) (*43*).

The formation of stable Eigen and Zundel cations is expected to be general on other noble metal surfaces with different reactivity. We have observed similar Eigen and Zundel structures on Ru(0001) and Cu(111) surfaces (figs. S25 and S26). Our PIMD simulations showed that the Zundel and Eigen structures observed in this work remained stable even up to room temperature despite thermal broadening (fig. S13). Further considering that the water layers near the liquid-solid interface can be relatively ordered (ice-like) because of the prominent metal-water interaction (*44*), especially when the electric potential is applied on the electrode (*45*), the low-temperature water monolayers on metal surfaces studied in this work may provide useful atomic-scale information to understand various electrode processes in aqueous environments. The coupled Eigen-Zundel interconversion and interfacial proton transfer are beyond the known elementary steps for H_2 production, which may help to boost the HER efficiency from a new perspective.

REFERENCES AND NOTES

1. D. Marx, M. E. Tuckerman, J. Hutter, M. Parrinello, *Nature* **397**, 601–604 (1999).
2. N. Agmon *et al*., *Chem. Rev.* **116**, 7642–7672 (2016).
3. K.-D. Kreuer, *Chem. Mater.* **8**, 610–641 (1996).
4. C. A. Wraight, *Biochim. Biophys. Acta Bioenerg.* **1757**, 886–912 (2006).
5. G. A. Voth, *Acc. Chem. Res.* **39**, 143–150 (2006).
6. M. Chen *et al*., *Nat. Chem.* **10**, 413–419 (2018).
7. S. J. Peighambardoust, S. Rowshanzamir, M. Amjadi, *Int. J. Hydrogen Energy* **35**, 9349–9384 (2010).
8. K.-D. Kreuer, S. J. Paddison, E. Spohr, M. Schuster, *Chem. Rev.* **104**, 4637–4678 (2004).
9. D. Marx, *ChemPhysChem* **7**, 1848–1870 (2006).
10. O. F. Mohammed, D. Pines, J. Dreyer, E. Pines, E. T. J. Nibbering, *Science* **310**, 83–86 (2005).
11. M. Garcia-Viloca, J. Gao, M. Karplus, D. G. Truhlar, *Science* **303**, 186–195 (2004).
12. I. S. Ramsey, M. M. Moran, J. A. Chong, D. E. Clapham, *Nature* **440**, 1213–1216 (2006).
13. M. Sasaki, M. Takagi, Y. Okamura, *Science* **312**, 589–592 (2006).
14. T. E. Springer, T. A. Zawodzinski, S. Gottesfeld, *J. Electrochem. Soc.* **138**, 2334–2342 (1991).
15. A. J. Bard, L. R. Faulkner, *Electrochemical Methods: Fundamentals and Applications* (Wiley, ed. 2, 2010).
16. W. Schmickler, E. Santos, *Interfacial Electrochemistry* (Springer, ed. 2, 2010).
17. Y. Kim, S. Shin, H. Kang, *Angew. Chem. Int. Ed.* **54**, 7626–7630 (2015).
18. G. Zundel, H. Metzger, *Z. Phys. Chem.* **58**, 225–245 (1968).
19. E. Wicke, M. Eigen, T. Ackermann, Eds., *Z. Phys. Chem* **1**, 340–364 (1954).
20. M. Thämer, L. De Marco, K. Ramasesha, A. Mandal, A. Tokmakoff, *Science* **350**, 78–82 (2015).
21. F. Dahms, B. P. Fingerhut, E. T. J. Nibbering, E. Pines, T. Elsaesser, *Science* **357**, 491–495 (2017).
22. J. A. Fournier, W. B. Carpenter, N. H. C. Lewis, A. Tokmakoff, *Nat. Chem.* **10**, 932–937 (2018).
23. C. Radüge, V. Pflumio, Y. R. Shen, *Chem. Phys. Lett.* **274**, 140–144 (1997).
24. S. Das, M. Bonn, E. H. G. Backus, *Angew. Chem. Int. Ed.* **58**, 15636–15639 (2019).
25. K. R. Asmis *et al*., *Science* **299**, 1375–1377 (2003).
26. J. M. Headrick *et al*., *Science* **308**, 1765–1769 (2005).
27. J. W. Shin *et al*., *Science* **304**, 1137–1140 (2004).
28. F. J. Giessibl, *Rev. Sci. Instrum.* **90**, 011101 (2019).
29. J. Peng *et al*., *Nat. Commun.* **9**, 122 (2018).
30. R. Ma *et al*., *Nature* **577**, 60–63 (2020).
31. J. Peng *et al*., *Nature* **557**, 701–705 (2018).
32. M. Wagner, B. Meyer, M. Setvin, M. Schmid, U. Diebold, *Nature* **592**, 722–725 (2021).
33. L. Gross, F. Mohn, N. Moll, P. Liljeroth, G. Meyer, *Science* **325**, 1110–1114 (2009).
34. P. Hapala *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **90**, 085421 (2014).
35. E. Skúlason *et al*., *Phys. Chem. Chem. Phys.* **9**, 3241–3250 (2007).
36. N. Bjerrum, *Science* **115**, 385–390 (1952).
37. T. Erdey-Grúz, M. Volmer, *Z. Phys. Chem.* **150A**, 203–213 (1930).
38. J. Heyrovský, *Recl. Trav. Chim. Pays Bas* **46**, 582–585 (1927).
39. A. Glebov, A. P. Graham, A. Menzel, J. P. Toennies, *J. Chem. Phys.* **106**, 9382–9385 (1997).
40. S. Nie, P. J. Feibelman, N. C. Bartelt, K. Thürmer, *Phys. Rev. Lett.* **105**, 026102 (2010).
41. J. Tafel, *Z. Phys. Chem.* **50U**, 641–712 (1905).
42. P. S. Rice, Z. P. Liu, P. Hu, *J. Phys. Chem. Lett.* **12**, 10637–10645 (2021).
43. N. M. Marković, B. N. Grgur, P. N. Ross, *J. Phys. Chem. B* **101**, 5405–5413 (1997).
44. T. Fukuma, Y. Ueda, S. Yoshioka, H. Asakawa, *Phys. Rev. Lett.* **104**, 016101 (2010).
45. C. Y. Li *et al*., *Nat. Mater.* **18**, 697–701 (2019).
46. Y. Tian *et al*., Zenodo (2022).

ACKNOWLEDGMENTS

The authors thank the computational resources provided by the TianHe-1A, TianHe II supercomputer, High-performance Computing Platform of Peking University. **Funding:** This work was supported by the National Key R&D Program under grants 2021YFA1400500, 2021YFA1400501, and 2017YFA0205003; the National Natural Science Foundation of China under grants 11888101, 11634001, 21725302, 21902013, 91961118, 11934003, and 11974024; the Strategic Priority Research Program of Chinese Academy of Sciences under grants XDB28000000 and XDB33000000; the Fundamental Research Funds for the Central Universities, Beijing Municipal Science and Technology Commission under grant Z181100004218006; the Beijing Natural Science Foundation under grant 1202016; the National Postdoctoral Program for Innovative Talents under grant BX2021040; and the China Postdoctoral Science Foundation under grant 2021M690408. **Author contributions:** Conceptualization and supervision: Y.J. and E.-G.W.; STM/AFM measurements: Y.T., J.H., J.G., S.Y., Z.W., D.G., X.L., Z.Z., and B.C.; Ab initio DFT calculations and AFM simulations: D.C., Y.S., and L.-M.X.; PIMD simulations: J.C. and X.-Z.L.; Interpretation of the data: Y.J., E.-G.W., J.C., J.G., L.-M.X., X.-Z.L, Y.T., D.C., J.H., Y.S., S.Y., Z.W., D.G., X.L., Z.Z., and B.C.; Writing and review: Y.J., J.C., J.G., Y.T., D.C., J.H., Y.S., S.Y., E.-G.W., L.-M.X., and X.-Z.L. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The tabulated data used to create the Figures for the main text have been deposited in Zenodo (*46*). All other data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo0823
Materials and Methods
Supplementary Text
Figs. S1 to S26
Tables S1 to S3
References (*47*–*74*)

Submitted 12 January 2022; accepted 16 May 2022
10.1126/science.abo0823

COSMOCHEMISTRY

Krypton in the Chassigny meteorite shows Mars accreted chondritic volatiles before nebular gases

Sandrine Péron*† and Sujoy Mukhopadhyay

Volatile elements are thought to have been delivered to Solar System terrestrial planets late in their formation through accretion of chondritic meteorites. Mars can provide information on inner Solar System volatile delivery during the earliest planet formation stages. We measured krypton isotopes in the martian meteorite Chassigny, representative of the planet's interior. We found chondritic krypton isotope ratios, which imply early incorporation of chondritic volatiles. The atmosphere of Mars has different (solar-type) krypton isotope ratios, indicating that it is not a product of magma ocean outgassing or fractionation of interior volatiles. Atmospheric krypton instead originates from accretion of solar nebula gas after formation of the mantle but before nebular dissipation. Our observations contradict the common hypothesis that during planet formation, chondritic volatile delivery occurred after solar gas acquisition.

Terrestrial planets acquire their volatile elements (e.g., hydrogen, carbon, nitrogen, noble gases) during formation. Models of this process often start with gases derived from the solar nebula (*1*), subsequently modified by fractionation during atmospheric escape and addition of volatiles from chondritic meteorites either during the main accretionary phase (*2*–*4*) or as a late veneer toward the end stages of planet formation (*1*, *5*). The exact sources and timing of these events remain under debate (*2*–*7*). For example, capture of nebular volatiles by planets is not universally accepted because chondritic signatures are observed for most volatile elements (*2*, *5*–*7*). Potential chondritic contributions come from accretion of solid bodies, originating from the inner Solar System with compositions similar to that of enstatite chondrite meteorites (*6*) or from the outer Solar System with compositions like that of carbonaceous chondrite meteorites or comets (*7*, *8*).

As a result of their chemical inertness, noble gases retain a record of volatile accretion and the physical processes associated with it. In Earth's mantle, helium and neon have solar-like isotope ratios whereas the heavy noble gases (argon, krypton, xenon) have isotope ratios that indicate a chondritic origin (*2*, *4*, *9*, *10*). These observations can be interpreted as either simultaneous accretion of solar and chondritic volatiles or early acquisition of solar volatiles followed by late addition of chondritic volatiles, provided the latter were mixed into Earth's interior (*2*, *11*). Because Mars was mostly formed during the first 4 million years (Myr) of Solar System formation (*12*), it can provide information on volatile accretion during the earliest planet formation stages.

The martian meteorite Chassigny contains trapped noble gases (*13*) thought to represent

Department of Earth and Planetary Sciences, University of California Davis, Davis, CA 95616, USA.
***Corresponding author. Email: sandrine.peron@erdw.ethz.ch**
†Present address: Institute of Geochemistry and Petrology, ETH Zürich, 8092 Zürich, Switzerland.

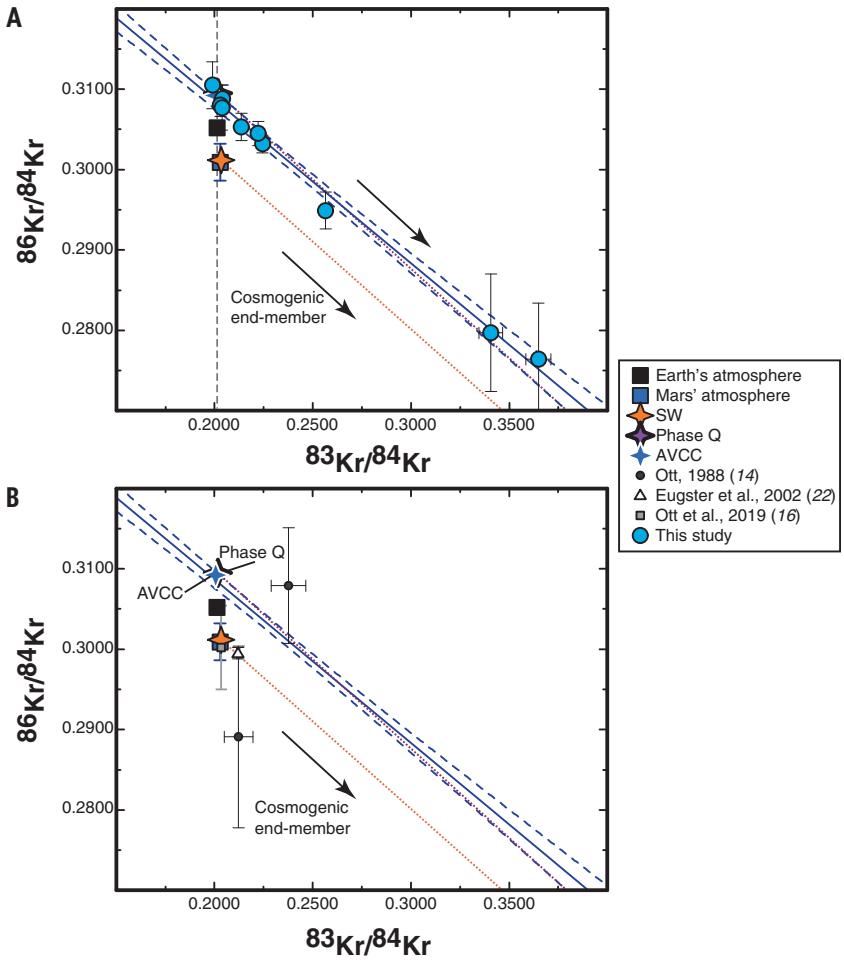


Fig. 1. Krypton isotope ratios for Chassigny. (**A**) Our measurements of Chassigny are shown as blue circles with each data point representing a different temperature step (because of contamination from Earth's air, the first four steps were discarded as mentioned in the text and so are not shown). The compositions of phase Q (*21*), AVCC (*13*), solar wind (SW) (*20*), and that of the atmospheres of Earth and Mars are shown with symbols indicated in the legend. The blue line is a linear model fitted to the Chassigny data with the 95% confidence interval shown as dashed lines. The dotted lines indicate values obtained by mixing phase Q (purple) and solar wind (orange) with cosmogenic krypton (*13*). The vertical dashed line indicates the $^{83}Kr/^{84}Kr$ ratio that is free of cosmogenic contributions (*13*). (**B**) Same data as (A) but compared with previous measurements of Chassigny (*14*, *16*, *22*). In both panels, error bars indicate 1σ uncertainties. Figure S1 shows additional details and isotope ratios.

interior composition of Mars (*14*–*16*). Analyses of xenon isotopes in Chassigny suggest that Mars' mantle contains solar-like xenon (*14*, *15*), and by inference, solar krypton (*16*). Xenon in Mars' atmosphere is mass-fractionated toward heavy isotopes but is consistent with an originally solar composition (*17*, *18*). Measurements taken from Mars rovers and martian meteorites have shown that the atmospheric krypton isotopic composition is indistinguishable from that of solar (*17*). Therefore, it has been argued that both the Martian mantle and atmosphere contain solar noble gases (*14*–*18*) with no indication of a chondritic contribution, potentially implying that Mars accreted all of its noble gases directly from the solar nebula.

However, xenon alone might not reliably indicate the sources of Mars' interior volatiles (*19*). The light isotopes (^{124}Xe, ^{126}Xe, ^{128}Xe, and ^{130}Xe)—which are stable and nonradiogenic—have nearly indistinguishable ratios for solar, chondritic, and cometary sources. The solar isotope ratios for the heavier isotopes (^{131}Xe, ^{132}Xe, ^{134}Xe, and ^{136}Xe) are intermediate between chondritic and cometary sources (*8*); further, these isotopes are also produced during spontaneous fission of ^{244}Pu (now extinct) and ^{238}U (still extant) (*10*).

Krypton isotopes in Chassigny could potentially discriminate between solar and chondritic sources as a result of their larger isotopic differences: Solar krypton is enriched in the light isotopes (relative to Earth's atmosphere) whereas chondritic krypton is enriched in the heavier isotopes (*20*, *21*). However, previous krypton isotopic measurements of Chassigny had insufficient precision to distinguish between solar and chondritic sources (*14*, *16*, *22*). In addition, Chassigny was exposed to cosmic rays during transit to Earth (11 Myr exposure age) (*14*), producing cosmogenic krypton isotopes from spallation reactions and partially masking the signature of trapped krypton (*14*).

We measured noble gas abundances and isotope ratios for Ne, Ar, Kr, and Xe in two separate samples of Chassigny with laser step-heating (temperature steps between 280° and 1570°C). We specifically developed a protocol for heavy noble gas separation and multicollector noble gas mass spectrometry (*13*). The krypton and xenon isotope ratios in Chassigny are shown in Figs. 1 and 2, respectively. Numerical values are listed in tables S1 and S2 and additional isotope combinations are plotted in figs. S1 to S3.

We find that the krypton data fall on a single line, reflecting mixing of cosmogenic gases with trapped Martian mantle gases (Fig. 1 and fig. S1). Except for ^{86}Kr, all Kr isotopes can be produced in spallation reactions with ^{83}Kr having the highest production rate (*13*). We therefore use the ^{86}Kr/^{84}Kr ratio to evaluate the source of Martian mantle heavy noble gases. By plotting ^{86}Kr/^{84}Kr as a function of ^{83}Kr/^{84}Kr, we determine the ^{86}Kr/^{84}Kr ratio corresponding to a ^{83}Kr/^{84}Kr value free of cosmogenic Kr; the result is the trapped mantle component (*13*).

Similarly, the xenon isotopic data fall on a mixing line (Fig. 2 and fig. S2) between cosmogenic and martian mantle compositions. The first four temperature steps between 280° and 575°C (tables S1 to S3) show a large contribution from Earth's air, with Ne, Ar, Kr, and Xe isotopic ratios either close to that of Earth's air or intermediate between air and the cosmogenic value (*13*). Because these are the initial low temperature steps they likely represent shallow contamination of Chassigny by Earth's atmosphere (*15*); further, because the subsequent heating steps do not show signs of Earth air contamination for neon, argon, krypton, or xenon we use them to infer the Martian mantle composition. We discard the first four steps in our subsequent analysis and discussion.

The ^{86}Kr/^{84}Kr ratio we measure for the interior of Mars differs from the solar composition but is indistinguishable from average carbonaceous chondrites (AVCC) (*13*). Figure 1 shows that a mixture of solar and cosmogenic Kr does not pass through any of the measured data points, ruling out solar Kr as the trapped component. The krypton isotopic data do fall on a mixing line between a chondritic and cosmogenic component (*13*). Selecting the ^{83}Kr/^{84}Kr ratio free of cosmogenic contributions, we find a martian mantle ^{86}Kr/^{84}Kr ratio of 0.3085 ± 0.0006 (1σ)—consistent with the chondritic value (table S1). The AVCC value seems to be a better match for the martian mantle composition than phase Q, a carbonaceous phase that carries most of the heavy noble gases in chondrites and sometimes appears as the only trapped composition in achondrites (carbonaceous chondrites have additional presolar components) (*21*). However, we cannot rule out a mixture of

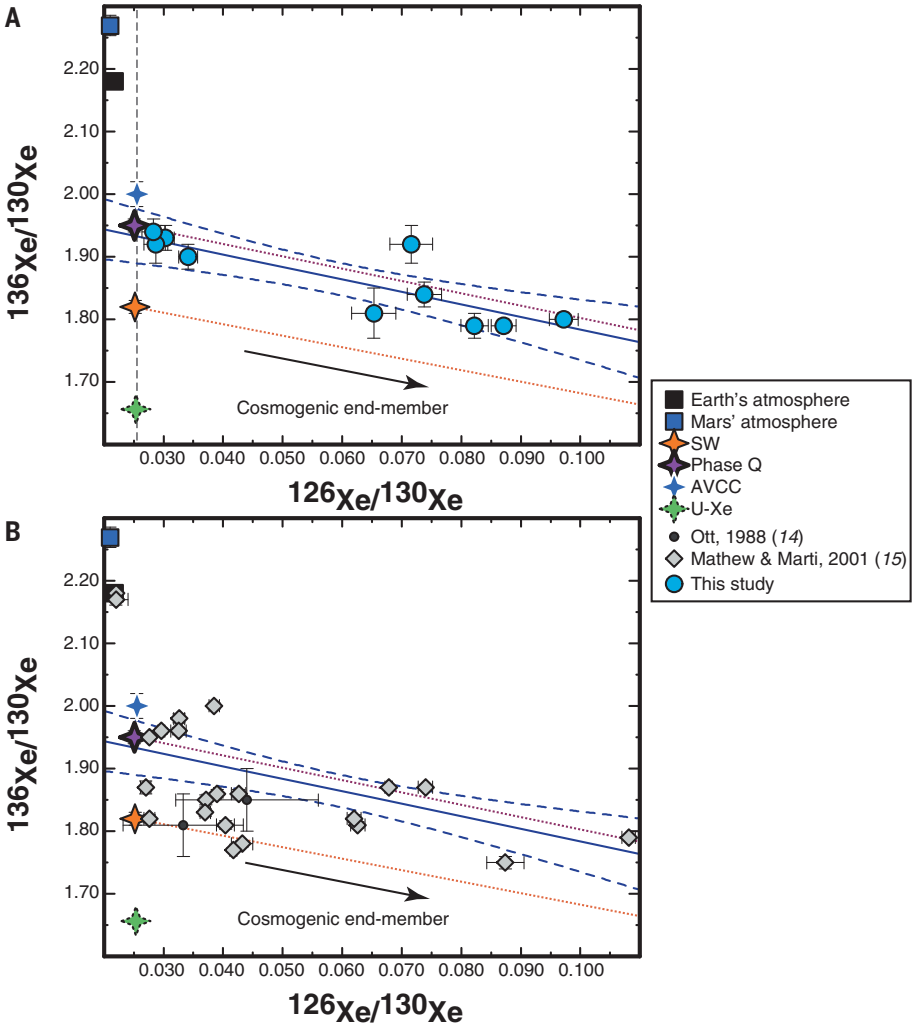


Fig. 2. Xenon isotope ratios for Chassigny. (**A** and **B**) Same as Fig. 1 but for xenon. Also shown is the U-Xe value, thought to be Earth's atmosphere precursor (*1*) and consisting of a mix of ~80% chondritic Xe with 20% cometary Xe (*8*).

phase Q gases with a small amount of solar gases as a potential match for the observed $^{86}Kr/^{84}Kr$ value. The similarity of the Mars mantle $^{86}Kr/^{84}Kr$ ratio to chondritic Kr cannot result from addition of fission-produced Kr to solar or cometary Kr (*13*). Previous Kr measurements of Chassigny (Fig. 1) precluded accurate determination of the martian mantle composition as these data either have large uncertainties (*14*, *16*) or consist of a single bulk measurement not targeted to determine the martian interior composition (*22*). The $^{86}Kr/^{84}Kr$ ratio we infer for the interior of Mars is closest to AVCC, so we conclude that chondritic gases were incorporated into the interior of Mars.

We find that chondritic noble gas ratios in the interior of Mars are consistent with the observed elemental abundance ratios—$^{36}Ar/^{132}Xe$ and $^{84}Kr/^{132}Xe$ (Fig. 3)—which are close to AVCC (*13*) and phase Q (*21*) values and consistent with previous results (*14*, *15*). Elemental abundance ratios can be modified by magma degassing and gas extraction in the laboratory (*15*), leading to variations in Fig. 3. However, all data points are close to the chondritic value and are distinct from the solar values as well as Mars and Earth atmosphere values.

Our xenon measurements are also consistent with chondritic gases in the interior of Mars. The $^{136}Xe/^{130}Xe$ ratio in Chassigny is distinct from the solar composition but close to the chondritic value and consistent with a single mixing line (Fig. 2). Extrapolating to an AVCC $^{126}Xe/^{130}Xe$ ratio of 0.0255 (*13*) yields a Mars mantle $^{136}Xe/^{130}Xe$ ratio of 1.933 ± 0.022 (1σ). However, as discussed above xenon isotopic compositions are more difficult to interpret as a result of multiple components, and the cosmic ray contributions for Xe are more difficult to correct (*13*). Nonetheless, our Xe data are consistent with a chondritic component though it is not required without incorporating the constraints from Kr.

Chondritic Kr and Xe in the martian interior do not preclude acquisition of other volatile species from the solar nebula. For example, the $^{15}N/^{14}N$ ratio of Chassigny could indicate solar-derived N, although enstatite chondrites might also be the source (*15*). Objects larger than one lunar mass can gravitationally capture an atmosphere from the solar nebula, which might then be incorporated into the solid body (*1*). Although a minor solar component cannot be ruled out, the lack of detectable solar Kr in Chassigny precludes incorporation of large amounts of solar Kr into the interior of Mars either through a magma ocean or by adsorption and burial beneath the surface during accretion (e.g., *1*).

Chondritic Kr in the martian mantle contrasts with solar Kr in the atmosphere of Mars (*17*, *18*). Cometary Kr—which is depleted in ^{83}Kr and ^{86}Kr relative to solar (*23*)—cannot account for the atmospheric Kr, suggesting that it was acquired from the solar nebula. Acquisition of the atmosphere from the solar nebula must have occurred after the interior incorporated chondritic Kr because otherwise chondritic Kr signatures would be seen in the atmosphere. Both interior and atmospheric gases were accreted before the nebular gas dissipated, on a time scale of ~4 Myr (*24*) and attributable to radiation from the early Sun. Hence, Mars formed quickly prior to complete nebular dissipation, accreting most of its mass and the solar atmosphere within 4 Myr after birth of the Solar System (*12*, *25*) (Fig. 4). The sequence of volatile accretion on Mars suggested by our data—chondritic followed by solar nebular—is the opposite of most models of planet formation in which chondritic volatile delivery follows solar gas acquisition (*1*).

The compositional differences between interior and atmosphere indicate that the atmosphere of Mars was not generated primarily through outgassing from its interior as is often assumed (*26*). Because the interior of Mars is enriched in heavier Kr isotopes compared with its atmosphere, outgassing followed by hydrodynamic loss is ruled out as that would leave the atmosphere enriched in heavier isotopes compared with the mantle. Delivery of chondritic volatiles to the surface of Mars after dissipation of the nebula was likely limited as it would have left a mixture of solar and chondritic signatures in the atmosphere. Although planetesimal impacts would contribute to the budget of rare nonvolatile elements (e.g., platinum group elements), they might not contribute substantially to Mars' volatile budget, particularly if the planetesimals were

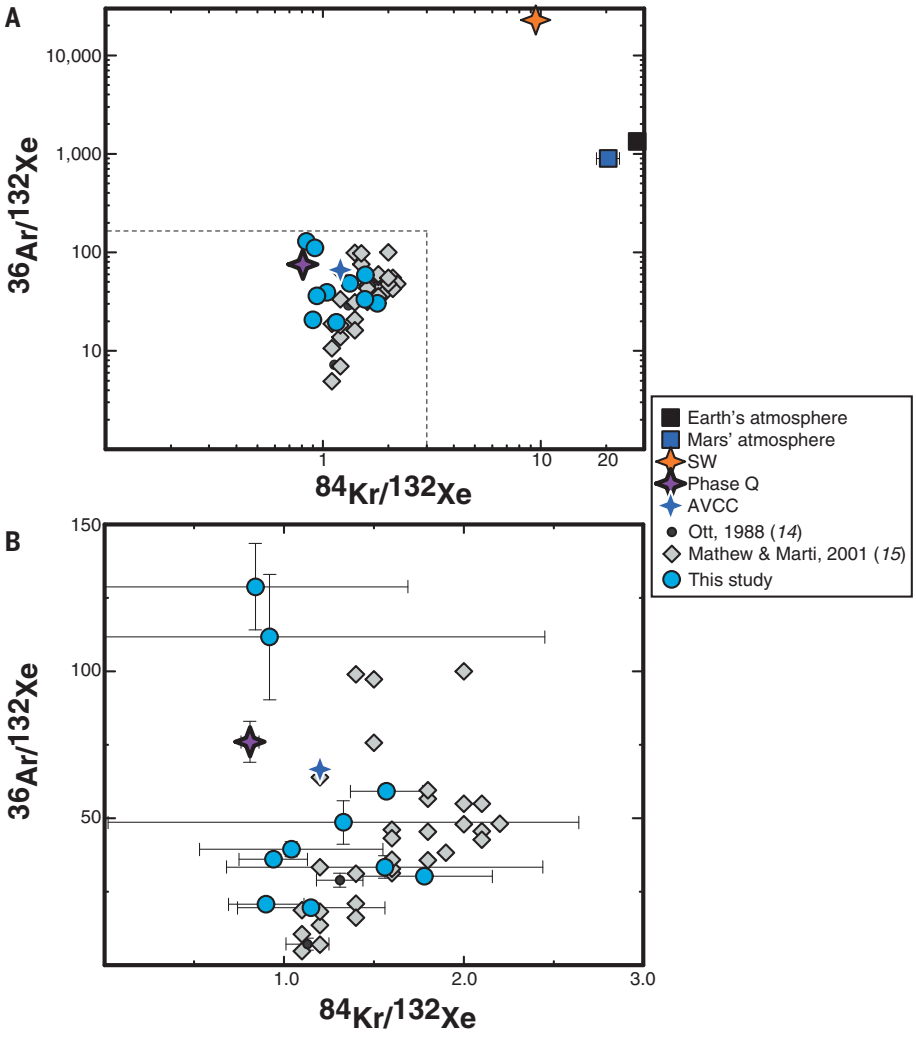


Fig. 3. Heavy noble gas abundances measured in Chassigny after correction for cosmogenic contributions. (**A**) Argon, krypton, and xenon element ratios from our measurements of Chassigny in each temperature step. For comparison, we show previous measurements of Chassigny (*14*, *15*), solar wind, phase Q, average carbonaceous chondrites, Earth's atmosphere, and martian atmosphere (references in table S4) with plotting symbols indicated in the legend. (**B**) Close up of the region within the dotted box in (A). Error bars show 1σ uncertainties and are smaller than the symbol size for solar wind and Earth's atmosphere.

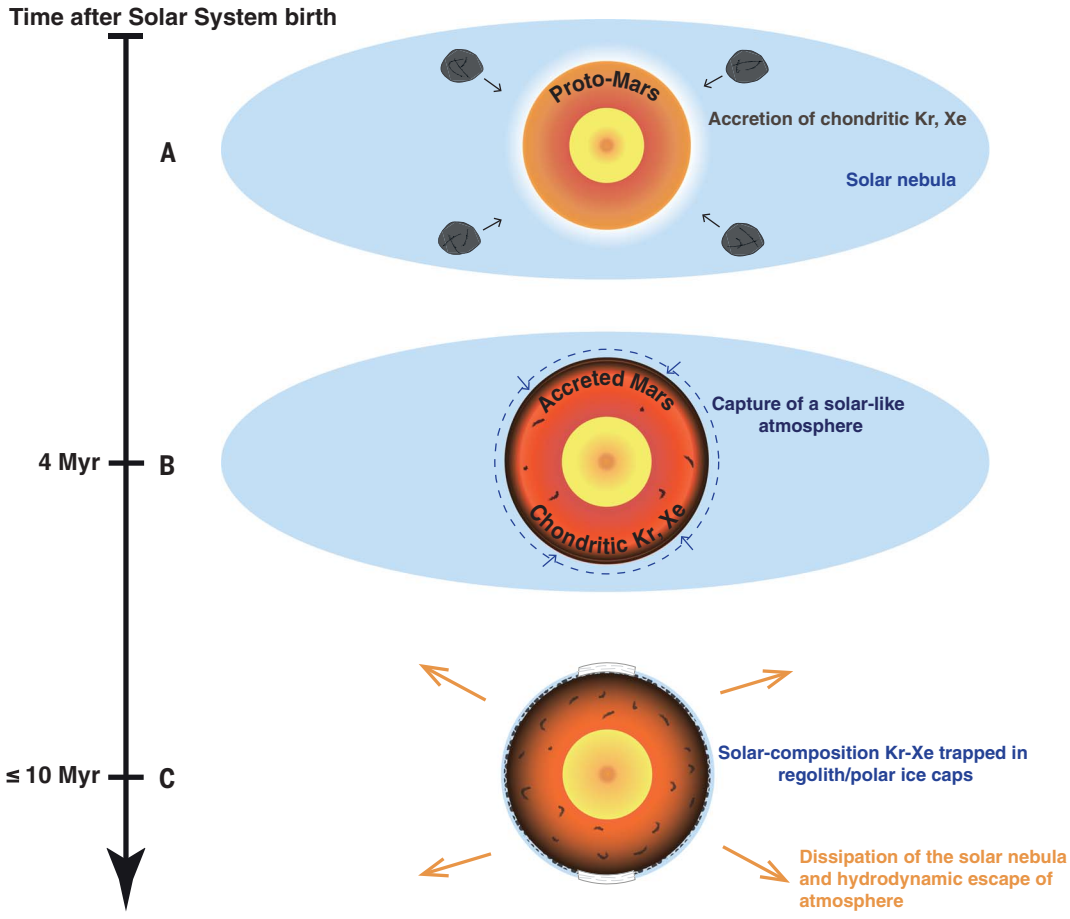


Fig. 4. Diagram illustrating a possible scenario for volatile delivery to Mars. (**A**) Accretion of chondritic Kr and Xe from planetesimals (gray), forming the mantle (orange) and core (yellow) of Mars. This occurs within the gaseous solar nebula (blue) during the first 1 to 4 Myr of Solar System formation. A magma ocean might have existed at this stage with either no atmosphere or a thin tenuous atmosphere (white grading into blue in the solar nebula). (**B**) After most of Mars' mass accreted (~4 Myr after Solar System birth) (*12*) and a solid lid formed on the surface, an atmosphere with solar isotope ratios is gravitationally captured (dashed blue circle). The atmosphere might have been relatively small, such that solar-composition gases were not incorporated into the interior in substantial quantities. (**C**) The surrounding nebula dissipates, halting accretion of solar-composition gases to atmospheric and surface reservoirs. Limited exchanges occurred between the heterogeneous mantle and the atmosphere. Atmospheric solar-composition Kr and Xe might have been trapped in polar ice caps (white) and/or in the sub-surface, preventing its loss during hydrodynamic escape of the atmosphere (orange arrows).

volatile-poor. Instead, planetesimal impacts could have eroded the atmosphere without inducing an isotopic fractionation, leading to net volatile loss (*27*).

A purely solar-like atmosphere would not persist on Mars if episodes of global magma ocean persisted well past the time of nebular dissipation and/or if there were episodes of hydrodynamic escape due to higher solar activity (*26*, *28*). A magma ocean would outgas and cause interior atmosphere exchange whereas hydrodynamic escape would cause the atmosphere to be lost or heavily fractionated (*26*, *28*). Mass fractionation of Xe from a solar precursor recorded in the Martian atmosphere (*17*) might not be a result of early hydrodynamic escape of neutral Xe, as Xe is the only noble gas that could escape as an ion in a photoionized hydrogen wind. This process has previously been invoked to explain the prolonged loss of Xe but not other noble gases during the Neoarchean era on Earth (*29*). Hence, Kr might be a better tracer of Mars' early atmospheric origin because it has kept its primordial solar signature. Gases lighter than Kr are lost from the modern martian atmosphere in a mass-dependent fractionation process due to solar wind bombardment (*30*).

If Mars captured its solar-like atmosphere from the nebula rather than acquiring it from mantle outgassing, it must have retained the solar composition Kr-Xe after the nebula dissipated. Thus, magma ocean phases on Mars ended before complete dissipation of the nebula, consistent with rapid mantle solidification in ~4 to 5 Myr (*28*, *31*). Hydrodynamic escape of the solar-like atmosphere is expected to be an efficient process on Mars after nebular dissipation (*26*, *28*). To prevent loss of the solar-composition Kr-Xe during hydrodynamic escape these gases might have been trapped in ice, either in the subsurface or in the polar ice caps (*32*). However, this scenario would require the surface of Mars to have remained cold—below the freezing point of water—after nebula dissipation. Later planetesimal impacts or episodic periods of warm climate would have released the trapped Kr-Xe into the atmosphere. Large-scale energetic events, such as large impacts that induce a magma ocean, are unlikely to have occurred after accretion of the interior and surface volatiles as they would have mixed the two reservoirs.

We conclude that within 4 Myr of Solar System formation, chondritic volatiles were incorporated into the interior of Mars in large quantities; heavy noble gases reach abundances up to two orders of magnitude higher than that in Earth's bulk mantle (*13*). The delivery of these chondritic volatiles to the inner Solar System could have been from material similar to enstatite chondrites (*6*) or from outer Solar System material scattered inward by giant planet migration (*33*).

REFERENCES AND NOTES

1. R. O. Pepin, D. Porcelli, *Rev. Mineral. Geochem.* **47**, 191–246 (2002).
2. M. W. Broadley *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **25**, 13997–14004 (2020).
3. B. Marty, *Earth Planet. Sci. Lett.* **313–314**, 56–66 (2012).
4. S. Péron, S. Mukhopadhyay, M. D. Kurz, D. W. Graham, *Nature* **600**, 462–467 (2021).
5. F. Albarède, *Nature* **461**, 1227–1233 (2009).
6. L. Piani *et al.*, *Science* **369**, 1110–1113 (2020).
7. C. M. O. Alexander *et al.*, *Science* **337**, 721–723 (2012).
8. B. Marty *et al.*, *Science* **356**, 1069–1072 (2017).
9. G. Holland, M. Cassidy, C. J. Ballentine, *Science* **326**, 1522–1525 (2009).
10. S. Péron, M. Moreira, *Geochem. Perspect. Lett.* **9**, 21–25 (2018).
11. C. L. Harper Jr., S. B. Jacobsen, *Science* **273**, 1814–1818 (1996).
12. N. Dauphas, A. Pourmand, *Nature* **473**, 489–492 (2011).
13. Materials and methods are available as supplementary materials.
14. U. Ott, *Geochim. Cosmochim. Acta* **52**, 1937–1948 (1988).

15. K. J. Mathew, K. Marti, *J. Geophys. Res.* **106** (E1), 1401–1422 (2001).
16. U. Ott, T. D. Swindle, S. P. Schwenzer, in *Volatiles in the Martian Crust*, J. Filiberto, S. P. Schwenzer, Eds. (Elsevier, 2019), pp. 35–70.
17. P. G. Conrad *et al*., *Earth Planet. Sci. Lett.* **454**, 1–9 (2016).
18. T. D. Swindle, *Rev. Mineral. Geochem.* **47**, 171–190 (2002).
19. M. Moreira, *Geochem. Perspect.* **2**, 229–403 (2013).
20. A. Meshik, C. Hohenberg, O. Pravdivtseva, D. Burnett, *Geochim. Cosmochim. Acta* **127**, 326–347 (2014).
21. H. Busemann, H. Baur, R. Wieler, *Meteorit. Planet. Sci.* **35**, 949–973 (2000).
22. O. Eugster, H. Busemann, S. Lorenzetti, D. Terribilini, *Meteorit. Planet. Sci.* **37**, 1345–1360 (2002).
23. M. Rubin *et al*., *Sci. Adv.* **4**, eaar6297 (2018).
24. H. Wang *et al*., *Science* **355**, 623–627 (2017).
25. R. Morishima, G. J. Golabek, H. Samuel, *Earth Planet. Sci. Lett.* **366**, 6–16 (2013).
26. H. Lammer *et al*., *Astron. Astrophys. Rev.* **26**, 2 (2018).
27. H. E. Schlichting, S. Mukhopadhyay, *Space Sci. Rev.* **214**, 34 (2018).
28. N. V. Erkaev *et al*., *Planet. Space Sci.* **98**, 106–119 (2014).
29. K. J. Zahnle, M. Gacesa, D. C. Catling, *Geochim. Cosmochim. Acta* **244**, 56–85 (2019).
30. B. M. Jakosky *et al*., *Science* **355**, 1408–1410 (2017).
31. L. T. Elkins-Tanton, *Earth Planet. Sci. Lett.* **271**, 181–191 (2008).
32. K. J. Zahnle, *J. Geophys. Res.* **98**, 10899–10913 (1993).
33. D. P. O'Brien, K. J. Walsh, A. Morbidelli, S. N. Raymond, A. M. Mandell, *Icarus* **239**, 74–84 (2014).

ACKNOWLEDGMENTS

We thank R. Wieler and two other anonymous reviewers for comments that greatly improved the manuscript. We thank the French National Museum of Natural History (MNHN, Paris, France) for graciously providing the Chassigny meteorite samples. We also thank D. Spaulding for assistance in preparing the samples. M. Huyskens, P.J. Carter, and S. Stewart are thanked for fruitful discussions and comments. **Funding:** Funds were provided by University of California, Davis, to S.M. **Author contributions:** S.P. and S.M. both designed the study, performed the noble gas analyses, interpreted the results, and wrote the manuscript. **Competing interests:** We declare no competing interests. **Data and materials availability:** Our samples of Chassigny were obtained from MNHN, 2525 PE2 and PE3. The samples were consumed during the experiments. Our noble gas (neon, argon, krypton, xenon) measurements are reported in tables S1 to S3.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk1175
Materials and Methods
Supplementary Text
Figs. S1 to S4
Tables S1 to S4
References (*34*–*69*)

View/request a protocol for this paper from *Bio-protocol*.

Submitted 23 June 2021; accepted 27 May 2022
10.1126/science.abk1175

NEUROSCIENCE

Neurogliaform cells dynamically decouple neuronal synchrony between brain areas

Ece Sakalar, Thomas Klausberger, Bálint Lasztóczi*

Effective communication across brain areas requires distributed neuronal networks to dynamically synchronize or decouple their ongoing activity. GABA*ergic* interneurons lock ensembles to network oscillations, but there remain questions regarding how synchrony is actively disengaged to allow for new communication partners. We recorded the activity of identified interneurons in the CA1 hippocampus of awake mice. Neurogliaform cells (NGFCs)—which provide GABA*ergic* inhibition to distal dendrites of pyramidal cells—strongly coupled their firing to those gamma oscillations synchronizing local networks with cortical inputs. Rather than strengthening such synchrony, action potentials of NGFCs decoupled pyramidal cell activity from cortical gamma oscillations but did not reduce their firing nor affect local oscillations. Thus, NGFCs regulate information transfer by temporarily disengaging the synchrony without decreasing the activity of communicating networks.

The brain is a complex system of networks interacting through concerted activity patterns broadcast through intricately structured connections (*1*, *2*). Rhythmic activation of neuronal assemblies in 10- to-30–ms time windows facilitates parsing of information by reader networks and generates transient gamma frequency (30 to 150 Hz) local field potential (LFP) oscillations (*3*–*5*). Gamma oscillations allow dynamic information routing (*6*, *7*) and neuronal circuits can perform active input selection if converging input pathways oscillate at different frequencies (*8*, *9*). However, many of the underlying brain mechanisms and network substrates remain unknown. In the hippocampus, sensory and mnemonic information from the entorhinal cortex and the CA3 area converge in the CA1 area (*10*) in which coordinated synaptic activity in terminals of temporoammonic (cortical) and Schaffer collateral (CA3) pathways give rise to mid-frequency ($gamma_M$; 75 Hz) and slow gamma oscillations ($gamma_S$; 37 Hz) in strata lacunosum-moleculare and radiatum, respectively (*11*–*13*). The association between afferent pathways and gamma oscillations paralleled by layer-specific arborizations of γ aminobutyric acid–expressing (GABA*ergic*) interneuron types make the rodent CA1 area a good candidate to explore input selection mechanisms (*14*, *15*).

We reasoned that activity of CA1 cells regulating cortico-hippocampal information flow would follow the dynamics of temporoammonic pathway that manifests as $gamma_M$ (*6*, *11*–*13*). To discover such neurons we simultaneously recorded layer-dependent gamma oscillations and neuronal spike timing in the dorsal hippocampal CA1 area of head-restrained mice running in a virtual corridor for a water reward (fig. S1) (*16*). To study $gamma_M$, $gamma_S$, and locally generated fast gamma oscillations ($gamma_F$; 120 Hz) (*12*) in isolation, volume-conducted LFP components were suppressed by calculating current source density (CSD; fig. S2) (*5*, *13*, *16*). Spike timing of most (84%) GABA*ergic* cells in CA1 ($n = 336$ cells) depended only weakly or not at all ($r < 0.07$) on the phase of $gamma_M$. However, a small neuron population (7.4%)—almost entirely located in stratum lacunosum-moleculare (23 of 25 cells)—showed distinctively strong phase locking ($r > 0.14$; Fig. 1 and fig. S3, supplementary text). Stratum lacunosum-moleculare also contained cells with little ($n = 22$) or no ($n = 28$) modulation by $gamma_M$ (Fig. 1C and fig. S3E). To identify the cells that fire phase locked to $gamma_M$ we labeled recorded cells with neurobiotin for post hoc histological analysis (*16*). Out of six successfully labeled stratum lacunosum-moleculare neurons, five showed strong coupling to $gamma_M$ with phase preference indistinguishable from other strongly coupled cells of this layer (Fig. 1, B to D; $P = 0.8$, Watson-Williams test; $n = 18$ cells). All five cells were identified as neurogliaform cells (NGFCs) (Fig. 1A and table S1, supplementary text). The spike timing of the sixth neuron was independent of $gamma_M$ ($P = 0.31$, Rayleigh test; $n = 297$ spikes), and this cell was not a NGFC (fig. S4; tables S1 and S2). Thus the population of GABA*ergic* stratum lacunosum-moleculare neurons with strong ($r = 0.27 \pm 0.07$) preferential firing on $gamma_M$ troughs ($\mu = 5.3 \pm 17.1°$; $n = 23$) corresponds to NGFCs (fig. S5; supplementary text). Firing of NGFCs was not coupled to $gamma_F$ and showed variable phase modulation by $gamma_S$ (figs. S6 and S7 and table S2). Oriens lacunosum-moleculare (OLM) cells also provide GABA*ergic* innervation to stratum lacunosum-moleculare but their soma and dendrites are located in stratum oriens (*14*). Spike timing of OLM cells was independent of $gamma_M$ but was moderately modulated by $gamma_F$ (fig. S8).

In awake rodents, 5 to 12 Hz theta oscillations occur during movement and irregular activity with intermittent sharp-wave ripple complexes (SWR) prevails during rest (Fig. 2A). The occurrence of SWRs had no effect on NGFC firing rate (fig. S9) (*17*), which markedly increased during theta oscillations (from 3.4 ± 3.9 Hz to 7.4 ± 4.7 Hz, $P = 4.6 \times 10^{-5}$, Wilcoxon signed-rank test; $n = 23$ cells; Fig. 2, A and B). Firing of some putative pyramidal cells (place cells) was restricted to sections of the corridor (place fields) and was phase precessing from ascending phase to peak of theta during traversals (Fig. 2 and fig. S10) (*18*). By contrast, NGFCs ($n = 16$, 2 identified and 14 putative) showed minimal spatial selectivity and constant theta phase preference (Fig. 2, A, C, and D, and fig. S10). Consequently, place cell spikes coincided with NGFC firing mostly on theta peaks upon place field exit (Fig. 2D and fig. S10A), when place cell firing is maximally modulated by $gamma_M$ (*13*, *19*).

Multisite recordings along the transverse axis of CA1 (fig. S11) disclosed widespread, tight, zero-lag phase synchrony and more spatially restricted amplitude correlations of $gamma_M$ (fig. S12, supplementary text). NGFCs fired on peaks of theta cycles ($r = 0.54 \pm 0.15$; $\mu = 206 \pm 18°$; $n = 23$) (*17*), coincident with high-amplitude $gamma_M$ ($r = 0.19 \pm 0.04$; $\mu = 193 \pm 8°$; $n = 63$ experiments) (*12*, *13*, *19*) implicating temporoammonic pathway $gamma_M$ synchrony in NGFC recruitment (Fig. 3A, fig. S13, and table S2). Indeed, within theta cycles NGFCs started to fire in high-amplitude $gamma_M$ cycles [Fig. 3B; $P = 6.8 \times 10^{-22}$; repeated measures one-way analysis of variance (ANOVA); $n = 23$]. To understand the consequences of NGFC activation we simulated the inhibitory postsynaptic $GABA_A$ conductance trace (g_{syn}) for NGFC spike trains (fig. S14). Because of its slow kinetics (*20*, *21*) NGFC-driven $GABA_A$ receptor-dependent inhibition may last for several $gamma_M$ cycles after the spike (fig. S14). This inhibition did not desynchronize $gamma_M$ per se as $gamma_M$ amplitude remained elevated after NGFC firing commenced (Fig. 3B). NGFCs may regulate cortico-hippocampal communication by releasing GABA onto apical dendritic tufts of CA1 pyramidal cells. In theta cycles, $gamma_M$ phase modulation of pyramidal cells first strengthened with the increasing $gamma_M$ amplitude (Fig. 3A and fig. S15A; $P = 2 \times 10^{-29}$; ANOVA with Tukey-Kramer correction; $n = 32$ experiments) but peaked earlier ($r = 0.21 \pm 0.08$; $\mu = 167 \pm 25°$; $n = 32$), conspicuously dropping as NGFC-dependent inhibition emerged phase-shifted by a quarter theta cycle from NGFC firing ($\mu = 288 \pm 18°$; $n = 23$; Fig. 3A). During the buildup of NGFC-dependent inhibition, pyramidal cell firing ramped (Fig. 3A). Pyramidal cell silencing after NGFC activation was not indicated in analysis of either cross correlograms (fig. S15,

Division of Cognitive Neurobiology, Center for Brain Research, Medical University of Vienna, Vienna, Austria.
*Corresponding author. Email: balint.lasztoczi@meduniwien.ac.at

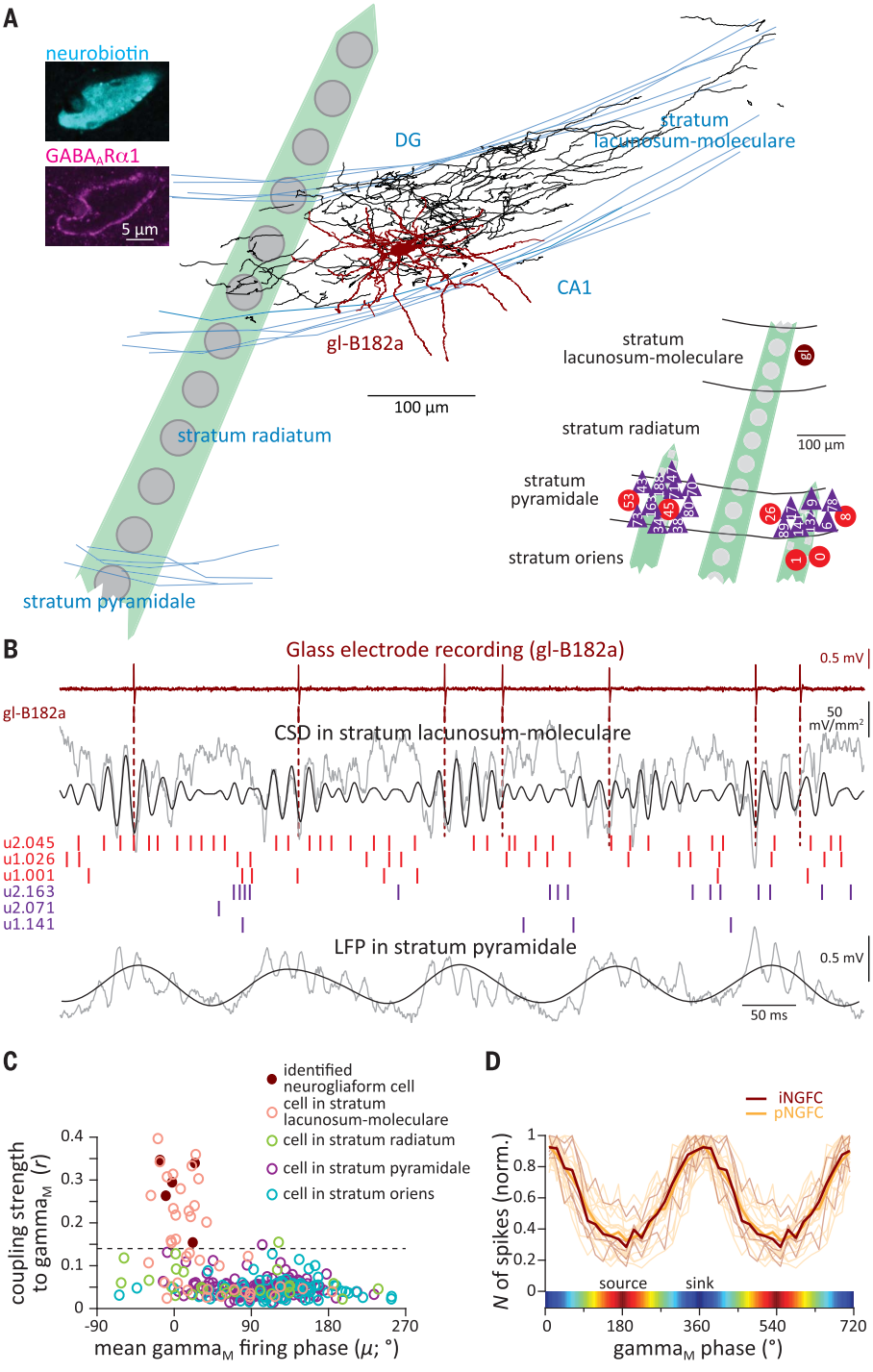


Fig. 1. NGFCs in the hippocampal CA1 area show distinctly strong coupling to mid-frequency gamma oscillations. (**A**) Reconstruction of somatodendritic (dark red) and axonal (black) arbors of cell gl-B182a recorded and labeled in experiment B182a and confocal scans showing the neurobiotin-labeled cell gl-ES9b (cyan) and GABA$_A$Rα1 subunit immunoreactivity (magenta). Cells gl-B182a and gl-ES9b were identified as NGFCs. (Inset) Schematic locations of putative GABA*ergic* (red circles) and pyramidal cells (purple triangles), and the cell gl-B182a recorded by the glass electrode (dark red circle) in experiment B182a. (**B**) Spike timing of neuron gl-B182a (dark red) and of other putative GABA*ergic* (red) and pyramidal cells (purple), together with gamma$_M$ (CSD in stratum lacunosum-moleculare, 53 to 90 Hz) and theta oscillations (LFP in stratum pyramidale, 5 to 12 Hz) in experiment B182a (gray, unfiltered traces). (**C**) Scatter plot of coupling strength versus preferred firing phase of all CA1 GABA*ergic* cells significantly coupled to gamma$_M$ (n = 262 cells). Estimated soma locations are color-coded; filled dark red circles represent identified NGFCs. Dotted line indicates threshold for strong coupling (0.14). (**D**) Distribution of spike counts of identified (iNGFC; dark red) and putative (pNGFC; orange) NGFCs as a function of gamma$_M$ phase (normalized to maxima; thick lines, mean).

C and D) or spike counts in gamma$_M$ cycles (Fig. 3B). To more directly probe the decoupling of CA1 from cortical inputs by NGFCs we compared phase coupling of pyramidal cell spikes in gamma$_M$ cycles before and after NGFC spikes. Immediately after the gamma$_M$ cycle hosting the first NGFC spike in a theta cycle, the coupling strength of pyramidal cell firing dropped (Fig. 3B and fig. S15E; $P = 1.1 \times 10^{-5}$; ANOVA with Tukey-Kramer correction; n = 14 experiments) and became largely not significant (fig. S16, A to C; α = 0.05; Rayleigh test), an effect specific to gamma$_M$ (fig. S16) and NGFCs (fig. S17). Although within gamma$_M$ cycles NGFCs fired 90° (3.3 ms) before pyramidal cells (Fig. 3C; n = 241 cells; r = 0.054 ± 0.021; μ = 98 ± 28° for pyramidal cells), in the cycle of the first NGFC spike the slow onset of inhibition permitted efficient cortico-hippocampal communication and therefore gamma$_M$ coupling of pyramidal cells remained elevated (Fig. 3B and fig. S17C). Decoupling was not a mere consequence of theta phase comodulation of NGFC firing, pyramidal cell gamma$_M$ coupling, and gamma$_M$ but instead depended on NGFC spike timing itself (figs. S18 and S19, supplementary text). Firing of putative GABA*ergic* cells in stratum pyramidale but not in stratum oriens also abruptly decoupled from gamma$_M$ oscillations after NGFC spikes (fig. S20). Thus, after NGFC activation the CA1 circuit decouples from cortical afferents (Fig. 3D).

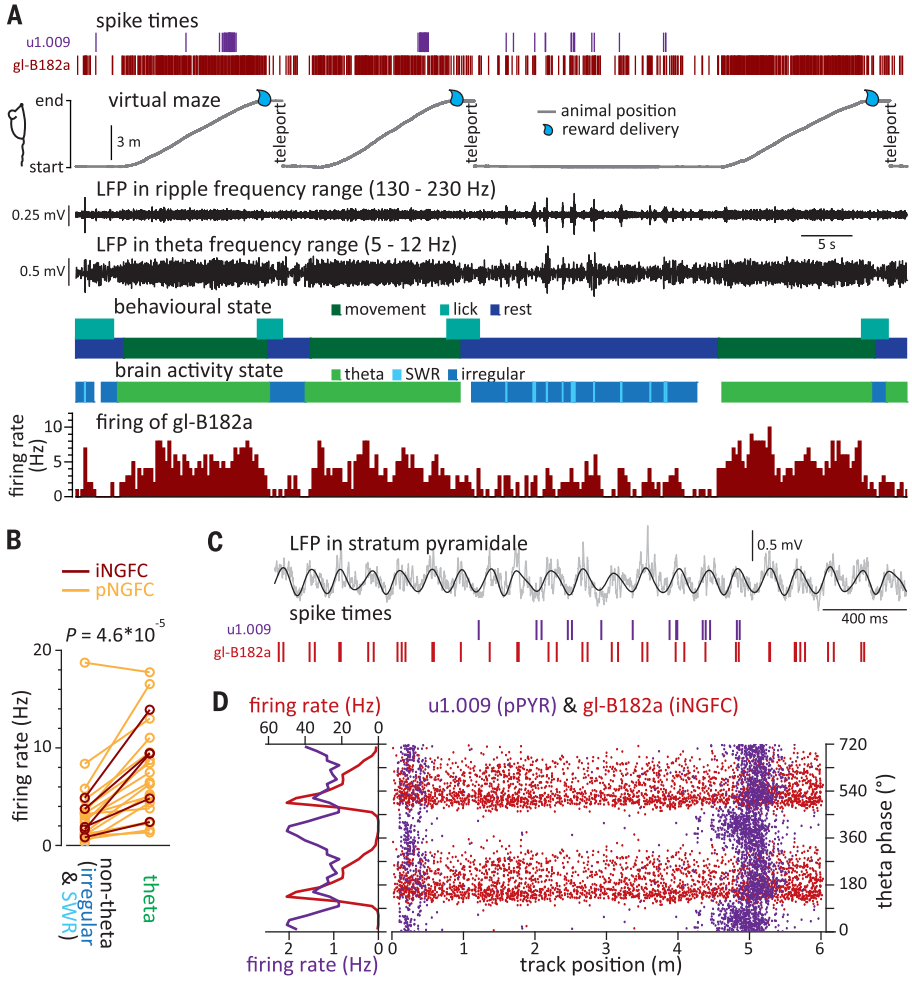


Fig. 2. NGFCs are activated during theta oscillations. (**A**) Spike times of an identified NGFC (gl-B182a; dark red) and putative pyramidal cell (u1.009, a place cell; purple) during three runs on the virtual corridor, together with ripple and theta oscillations from stratum pyramidale. Behavioral and brain states are indicated by color-coded boxes. Lower plot: Firing rate of cell gl-B182a in 300-ms windows. (**B**) Comparison of NGFC firing rates during theta versus nontheta periods. (**C**) Spike times of cells in (A) on an expanded time scale as the animal walks through the place field of u1.009, together with theta oscillations (5 to 12 Hz LFP in stratum pyramidale; gray, unfiltered trace). (**D**) Theta phase versus position raster plot of spikes of gl-B182a (dark red) and u1.009 (phase precessing place cell, purple). (Left) Theta phase dependence of the firing rates.

We discovered a network mechanism for dynamic regulation of cortico-hippocampal information transfer in the CA1 area. NGFCs release GABA to stratum lacunosum-moleculare, inducing slow inhibition in all pyramidal cell apical dendritic tufts within their axonal arbor (*20*–*22*). The faster $GABA_A$ component of this indiscriminate, layer-specific inhibition mediated by unitary volume transmission disconnects pyramidal cells from cortical afferents for a fraction of a theta cycle reported by a temporary decoupling of their spike timing from $gamma_M$ after NGFC firing. Summating over several theta cycles $GABA_B$ receptor mediated processes may regulate inputs on behavioral time scales (*14*, *20*, *22*). Cortical afferents contribute little to pyramidal cell firing rates but are indispensable for intact temporal organization and place fields in CA1 (*23*). This explains maintained pyramidal cell firing despite reduced cortico-hippocampal communication. The distal location of cortical synapses limits their influence (*24*) and therefore modulation of CA1 pyramidal cell firing by $gamma_M$ is generally weak (*12*, *13*). Cortico-hippocampal information transfer and coupling to $gamma_M$ can strengthen with cognitive load (*25*, *26*) during some network operations (*13*, *26*–*28*) and pathway interactions (*24*, *27*, *29*), implying dynamic control; inhibition by NGFCs provides a mechanism to exercise such control. Thalamic afferents also target NGFCs (*30*), further increasing the versatility of this cell type.

During exploration theta oscillations organize hippocampal activity, modulate gamma oscillation amplitudes (*11*–*13*, *31*), and segment pyramidal cell firing sequences (*32*). On theta peaks, activity and $gamma_M$ synchrony build up in the temporoammonic pathway and give rise to $gamma_M$ in stratum lacunosum-moleculare (*12*, *31*), which entrains CA1 pyramidal cell spikes when theta firing sequences start (*13*). This waxing rhythmic excitation also induces spiking in NGFCs with a lag (*22*), first when the amplitude of $gamma_M$ is already high. Late spiking and slowly rising postsynaptic currents of NGFCs ensure a window of efficient cortico-hippocampal information transfer before NGFCs detune pyramidal cells from $gamma_M$, allowing other pathways to control pyramidal cell recruitment to theta sequences (*13*). Thus NGFCs minimize input interference and optimize conditions for cooperative synaptic plasticity (*24*, *27*, *29*). Waning cortical excitation (*31*) and waxing inhibition from OLM cells (*33*–*35*) silence NGFCs on theta troughs, which prepares the network for the next cycle of $gamma_M$ synchronization through recovering the dynamic range of inhibition.

Neurogliaform cells are key regulators of cortical information flow to CA1, orchestrating precise integration of sensory and mnemonic information. Fast-spiking GABA*ergic* cells facilitate cortical communication by conducting gamma oscillations (*14*, *36*, *37*). By contrast, NGFCs with input layer-associated axonal and dendritic arbors, ubiquitous in cortical circuits (*22*, *38*, *39*), detune principal cell firing and regulate information flow by afferent-specific decoupling.

REFERENCES AND NOTES

1. S. W. Oh *et al.*, *Nature* **508**, 207–214 (2014).
2. A. Pouget, P. Dayan, R. Zemel, *Nat. Rev. Neurosci.* **1**, 125–132 (2000).
3. G. Buzsáki, *Neuron* **68**, 362–385 (2010).
4. P. Fries, *Neuron* **88**, 220–235 (2015).
5. B. Pesaran *et al.*, *Nat. Neurosci.* **21**, 903–919 (2018).

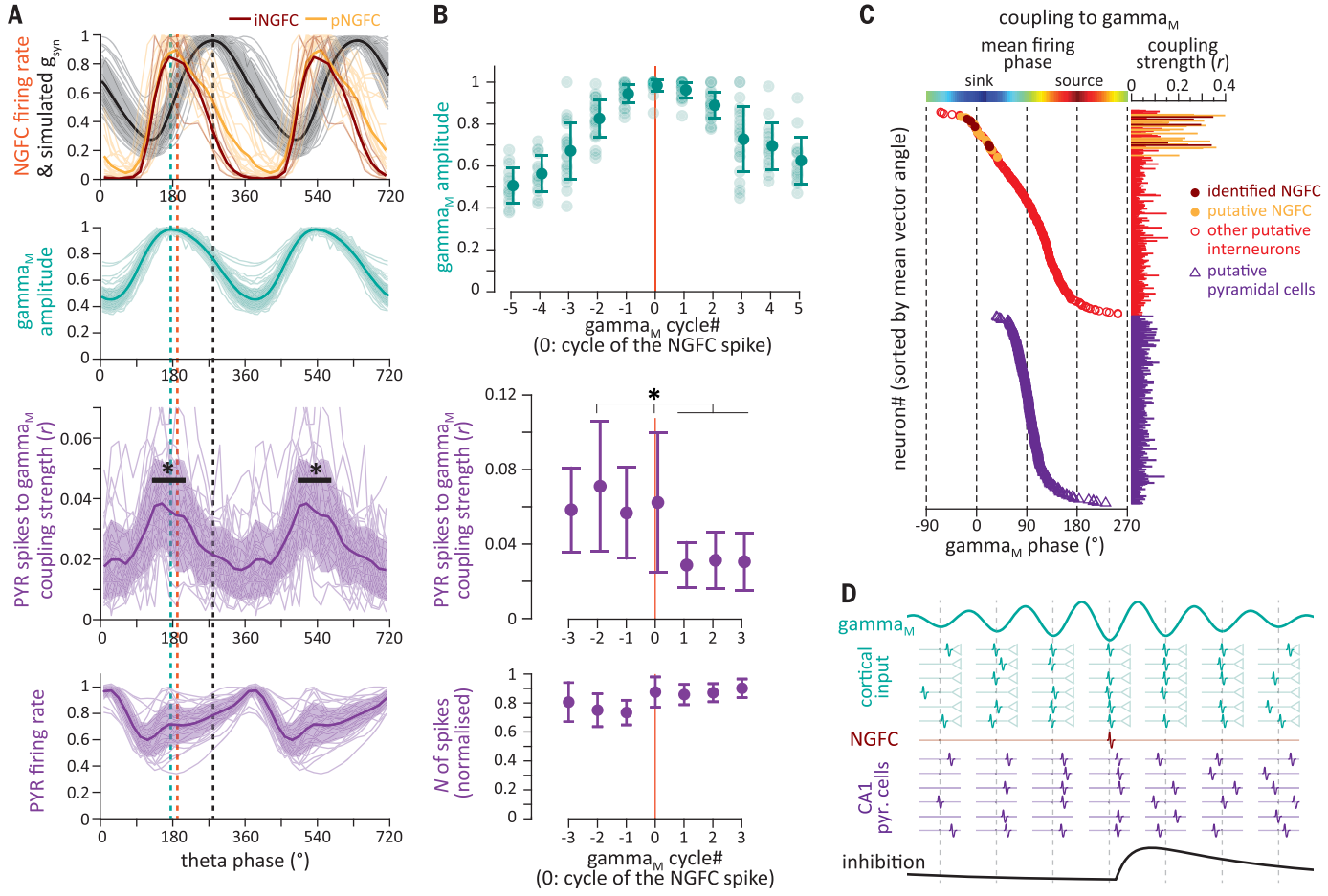


Fig. 3. Action potentials of NGFCs decouple pyramidal cell firing from mid-frequency gamma oscillations but do not suppress their activity. (**A**) Theta phase modulation of NGFC firing rate (dark red and orange for identified and putative NGFCs, respectively; normalized) and the simulated resultant inhibitory postsynaptic conductance in pyramidal cells (g_{syn}; black; normalized); of gamma$_M$ amplitude (normalized, turquoise); of coupling strength of pyramidal cell firing to gamma$_M$ (purple); and of pyramidal cell activity (purple, normalized). Light lines indicate individual experiments (pyramidal cells pooled); thick lines and shading indicate mean ± SD; asterisk, significant enhancement (α = 0.05, ANOVA with Tukey-Kramer correction). (**B**) Amplitude of gamma$_M$ (normalized, turquoise), and phase-coupling strength (r, purple, middle) and number (purple, bottom) of pyramidal cell spikes in gamma$_M$ cycles before (cycle number < 0), after (cycle number > 0) and during (cycle 0) the first NGFC spike in a theta cycle (asterisks indicate P < 0.05, ANOVA with Tukey-Kramer correction). (**C**) Mean firing phase (μ) and coupling strength (r) of identified (dark red) and putative (orange) NGFCs, other GABA*ergic* cells (red) and pyramidal cells (purple triangles) ordered by the mean firing phase (only cells significantly modulated by gamma$_M$ are plotted). (**D**) Schematic illustration of how NGFC activation affects cortico-hippocampal communication through gamma$_M$ oscillations.

6. T. Akam, D. M. Kullmann, *Neuron* **67**, 308–320 (2010).
7. G. G. Gregoriou, S. J. Gotts, H. Zhou, R. Desimone, *Science* **324**, 1207–1210 (2009).
8. A. Palmigiano, T. Geisel, F. Wolf, D. Battaglia, *Nat. Neurosci.* **20**, 1014–1022 (2017).
9. A. Fernández-Ruiz *et al*., *Science* **372**, eabf3119 (2021).
10. N. M. van Strien, N. L. Cappaert, M. P. Witter, *Nat. Rev. Neurosci.* **10**, 272–282 (2009).
11. L. L. Colgin *et al*., *Nature* **462**, 353–357 (2009).
12. E. W. Schomburg *et al*., *Neuron* **84**, 470–485 (2014).
13. B. Lasztóczi, T. Klausberger, *Neuron* **91**, 34–40 (2016).
14. K. A. Pelkey *et al*., *Physiol. Rev.* **97**, 1619–1747 (2017).
15. T. P. Vogels, L. F. Abbott, *Nat. Neurosci.* **12**, 483–491 (2009).
16. Materials and methods are available as supplementary materials.
17. P. Fuentealba *et al*., *J. Neurosci.* **30**, 1595–1609 (2010).
18. J. O'Keefe, M. L. Recce, *Hippocampus* **3**, 317–330 (1993).
19. M. Guardamagna, F. Stella, F. P. Battaglia, bioRxiv 2021.12.22.473863v1 [Preprint] (2021); https://doi.org/10.1101/2021.12.22.473863.
20. C. J. Price, R. Scott, D. A. Rusakov, M. Capogna, *J. Neurosci.* **28**, 6974–6982 (2008).
21. S. Oláh *et al*., *Nature* **461**, 1278–1281 (2009).
22. L. Overstreet-Wadiche, C. J. McBain, *Nat. Rev. Neurosci.* **16**, 458–468 (2015).
23. M. I. Schlesiger *et al*., *Nat. Neurosci.* **18**, 1123–1132 (2015).
24. T. Jarsky, A. Roxin, W. L. Kath, N. Spruston, *Nat. Neurosci.* **8**, 1667–1676 (2005).
25. V. Lopes-dos-Santos *et al*., *Neuron* **100**, 940–952.e7 (2018).
26. A. Fernández-Ruiz *et al*., *Neuron* **93**, 1213–1226.e5 (2017).
27. K. C. Bittner *et al*., *Nat. Neurosci.* **18**, 1133–1142 (2015).
28. H. O. Cabral *et al*., *Neuron* **81**, 402–415 (2014).
29. M. Remondes, E. M. Schuman, *Nature* **416**, 736–740 (2002).
30. R. Chittajallu *et al*., *Nat. Commun.* **8**, 152 (2017).
31. K. Mizuseki, A. Sirota, E. Pastalkova, G. Buzsáki, *Neuron* **64**, 267–280 (2009).
32. W. E. Skaggs, B. L. McNaughton, M. A. Wilson, C. A. Barnes, *Hippocampus* **6**, 149–172 (1996).
33. D. Elfant, B. Z. Pál, N. Emptage, M. Capogna, *Eur. J. Neurosci.* **27**, 104–113 (2008).
34. C. Varga, P. Golshani, I. Soltesz, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2726–E2734 (2012).
35. L. Katona *et al*., *Neuron* **82**, 872–886 (2014).
36. J. A. Cardin *et al*., *Nature* **459**, 663–667 (2009).
37. B. V. Atallah, M. Scanziani, *Neuron* **62**, 566–577 (2009).
38. B. Schuman *et al*., *J. Neurosci.* **39**, 125–139 (2019).
39. X. Jiang *et al*., *Science* **350**, aac9462 (2015).
40. E. Sakalar, Zenodo (2022); https://doi.org/10.5281/zenodo.6580307.

ACKNOWLEDGMENTS

We thank A. Wallerus for his help with creating the flat maps of the stratum lacunosum-moleculare (including creating and sharing software) and for his help with cell reconstruction and histological processing. The technical assistance of R. Hauer, M. S. Mateos, and E. Dögl is highly appreciated. We thank A. Wallerus, I. Vörösházy, and A. Leonhartsberger for training some of the animals and I. Vörösházy for performing some of the recordings. We thank P. Scholze for providing us with the experimental subjects and P. Scholze and W. Sieghart for their generous gift of antibodies against the α1 and δ subunits of the GABA$_A$ receptor. **Funding:** This work was supported by the FWF Austrian Science Fund, Stand-Alone Project P-29744-B27 (to B.L.). **Author contributions:** Conceptualization: B.L., E.S., and T.K. Methodology: B.L., E.S., and T.K. Investigation: E.S. and B.L. Formal analysis: E.S. and B.L. Software: B.L. and E.S. Data curation: E.S., B.L., and T.K. Visualization: E.S., B.L., and T.K.

Funding acquisition: B.L. and T.K. Project administration: B.L., T.K., and E.S. Supervision: B.L. and T.K. Writing – original draft: E.S. and B.L. Writing – review and editing: E.S., B.L., and T.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript and supplementary materials. Custom MATLAB scripts can be downloaded from (*40*). **License information:**

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo3355
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 and S2
References (*41*–*76*)

View/request a protocol for this paper from *Bio-protocol*.

Submitted 27 January 2022; accepted 30 May 2022
10.1126/science.abo3355

IMMUNOLOGY

ZAKα-driven ribotoxic stress response activates the human NLRP1 inflammasome

Kim S. Robinson[1,2], Gee Ann Toh[3], Pritisha Rozario[3], Rae Chua[3], Stefan Bauernfried[4,5], Zijin Sun[3], Muhammad Jasrie Firdaus[3], Shima Bayat[3], Rhea Nadkarni[6], Zhi Sheng Poh[3], Khek Chian Tham[2], Cassandra R. Harapas[7], Chrissie K. Lim[8]†, Werncui Chu[6], Celest W. S. Tay[3], Kiat Yi Tan[2], Tianyun Zhao[8], Carine Bonnard[1,2], Radoslaw Sobota[8], John E. Connolly[8], John Common[2], Seth L. Masters[7,9], Kaiwen W. Chen[10], Lena Ho[6,8], Bin Wu[11], Veit Hornung[4,5], Franklin L. Zhong[1,3]*

Human NLRP1 (NACHT, LRR, and PYD domain-containing protein 1) is an innate immune sensor predominantly expressed in the skin and airway epithelium. Here, we report that human NLRP1 senses the ultraviolet B (UVB)- and toxin-induced ribotoxic stress response (RSR). Biochemically, RSR leads to the direct hyperphosphorylation of a human-specific disordered linker region of NLRP1 ($NLRP1^{DR}$) by MAP3K20/ZAKα kinase and its downstream effector, p38. Mutating a single ZAKα phosphorylation site in $NLRP1^{DR}$ abrogates UVB- and ribotoxin-driven pyroptosis in human keratinocytes. Moreover, fusing $NLRP1^{DR}$ to CARD8, which is insensitive to RSR by itself, creates a minimal inflammasome sensor for UVB and ribotoxins. These results provide insight into UVB sensing by human skin keratinocytes, identify several ribotoxins as NLRP1 agonists, and establish inflammasome-driven pyroptosis as an integral component of the RSR.

The innate immune system uses germline-encoded sensor proteins to recognize conserved pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs). NACHT, LRR, and PYD domain-containing proteins (NLRPs) assemble the inflammasome complex in response to intracellular pathogens or stress signals, resulting in pyroptotic cell death characterized by caspase-1 activation, gasdermin D (GSDMD) pore formation, and interleukin 1 (IL-1) secretion (*1*–*3*). Human NLRP1 is notable among inflammasome sensors because of its unusual domain arrangement and divergence from rodent counterparts (*4*–*6*). Inhibitors of proteases DPP8 and DPP9, such as Val-boro-Pro (VbP) (*7*, *8*), are the only known molecules that can activate both rodent and human NLRP1 and a related human inflammasome sensor, caspase activation and recruitment domain-containing protein 8 (CARD8) (*9*–*11*). Recent work has shown that human NLRP1 senses double-stranded viral RNA, viral proteases (*12*–*14*), and ultraviolet B (UVB) irradiation (*15*). None of these triggers activates rodent NLRP1s, which in turn sense bacterial and protozoan toxins (*16*–*18*). All known NLRP1 triggers require the proteasomal degradation of the autoinhibitory N-terminal fragment (NT) and the oligomerization of the liberated C-terminal fragment (CT) to drive inflammasome activation (*12*, *19*–*21*).

Human NLRP1 is predominantly expressed in the skin and airway epithelia (*12*, *22*, *23*). Both rare germline mutations in *NLRP1* and common *NLRP1* single-nucleotide polymorphisms are associated with human skin disorders (*22*, *24*–*26*). Thus, human NLRP1 plays a unique role in skin immunity. UVB radiation with wavelengths of 280 to 315 nm, which is responsible for acute sunburn, is the most relevant to the skin of all NLRP1 triggers identified thus far (*15*). However, the molecular mechanisms by which NLRP1 senses UVB are unclear.

Using the N/TERT-1 immortalized human keratinocyte cell line (hereafter referred to as N/TERTs), we confirmed published findings (*15*, *27*) that UVB irradiation causes NLRP1-dependent pyroptosis in a dose-dependent manner as measured by IL-1β secretion, GSDMD cleavage, ASC oligomerization, and rapid propidium iodide (PI) uptake (within 4 to 6 hours) (fig. S1, A to C and E). UVB-induced IL-1β secretion required caspase-1 activity but not NLRP3 (fig. S1, B and D). Similar to other NLRP1 activators, UVB-dependent NLRP1 activation was accompanied by a decrease in NLRP1 NT and was blocked by the NEDD8/cullin inhibitor MLN4924 (figs. S1D and S2), which has been reported to regulate NLRP1-NT and CARD8-NT turnover (*12*, *28*). Treating N/TERT cells with toxic doses of the DNA-damaging chemicals camptothecin, etoposide, and cisplatin or hydrogen peroxide failed to induce NLRP1-dependent IL-1β secretion or PI uptake at early time points (fig. S3, A and B). Thus, neither DNA damage nor oxidative damage by free radicals alone is the primary driver of UVB-induced NLRP1 inflammasome activation (Fig. 1A).

We hypothesized that UVB-driven RNA damage activates the NLRP1 inflammasome. To test this, we pretreated cells with the nucleoside analog 4-thiouridine (4-SU), which selectively sensitizes RNA to ultraviolet A (UVA) radiation. Otherwise, UVA does not cause acute damage to unmodified nucleic acids because of its lower energy (*29*). Only in N/TERT cells pretreated with 4-SU did UVA cause the accumulation of thymine photoadducts in the cytoplasm, which was consistent with RNA damage (fig. S3D). 4-SU+UVA induced the secretion of IL-1β (Fig. 1B and fig. S3C) and phosphorylation of the stress-activated kinases (SAPKs) p38 and JNK, similar to the effects of UVB (Fig. 1B and fig. S3C). By contrast, UVA irradiation of N/TERT cells pretreated with a DNA sensitizer 5-bromo-2′-deoxyuridine did not induce IL-1β secretion and only effected weak SAPK phosphorylation (Fig. 1B and fig. S3C). Thus, RNA photodamage is more likely to be the upstream signal for UV-induced NLRP1 activation in keratinocytes.

Recently, the proximal sensor for UVB-triggered SAPK activation was found to be the long-splice isoform of the MAP3K20, also known as ZAKα kinase (*30*, *31*). ZAKα senses ribosomes that have stalled and/or collided after encountering a translocation-blocking mRNA lesion, such as those induced by UVB. Activated ZAKα undergoes extensive self-phosphorylation and phosphorylates downstream SAPKs. Collectively, this pathway is called the ribotoxic stress response (RSR) (Fig. 1A). UVB induced bona fide RSR activation in N/TERT keratinocytes marked by ZAKα, p38, and JNK phosphorylation at earlier time points and ZAKα degradation at later time points (Fig. 1B and fig. S4A). P38 and JNK phosphorylation was completely abrogated in CRISPR/Cas9 *ZAK* knockout (KO) N/TERT cells (Fig. 1B and fig. S4, B and C) after UVB or 4SU+UVA. *ZAK* KO or inhibition of ZAKα kinase activity by nilotinib blocked UVB-induced pyroptosis in N/TERT cells as measured by IL-1β secretion (Fig. 1, B and C, and fig. S4, D and E), GSDMD p30 cleavage (Fig. 1D), and rapid PI uptake (Fig. 1, E and F). Moreover, a specific ZAKα inhibitor, M443 (*32*), blocked UVB-induced IL-1β secretion in human skin explants (fig. S4, F and G). *ZAK* KO or ZAKα inhibition did not affect VbP-driven pyroptosis in N/TERT cells (figs. S4E and S5F). Thus, ZAKα is selectively required for the NLRP1 inflammasome activation downstream of UVB. In a 293T-ASC-green fluorescent protein (GFP) inflammasome reporter cell line (fig. S5, A and B),

[1]Skin Research Institute of Singapore (SRIS), 308232 Singapore. [2]Agency for Science, Technology and Research (A*STAR) Skin Research Laboratories (ASRL), 138648 Singapore. [3]Lee Kong Chian School of Medicine, Nanyang Technological University, 308232 Singapore. [4]Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany. [5]Max-Planck Institute of Biochemistry, 82152 Martinsried, Germany. [6]Cardiovascular Metabolic Disorders Program, Duke-NUS Medical School, 169857 Singapore. [7]Inflammation Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia. [8]Institute of Molecular and Cell Biology, A*STAR, 138673 Singapore. [9]Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010, Australia. [10]Immunology Translational Research Programme and Department of Microbiology & Immunology, Yong Loo Lin School of Medicine, National University of Singapore, 117545 Singapore. [11]School of Biological Sciences, Nanyang Technological University (NTU), 639798 Singapore.
*Corresponding author. Email: franklin.zhong@ntu.edu.sg

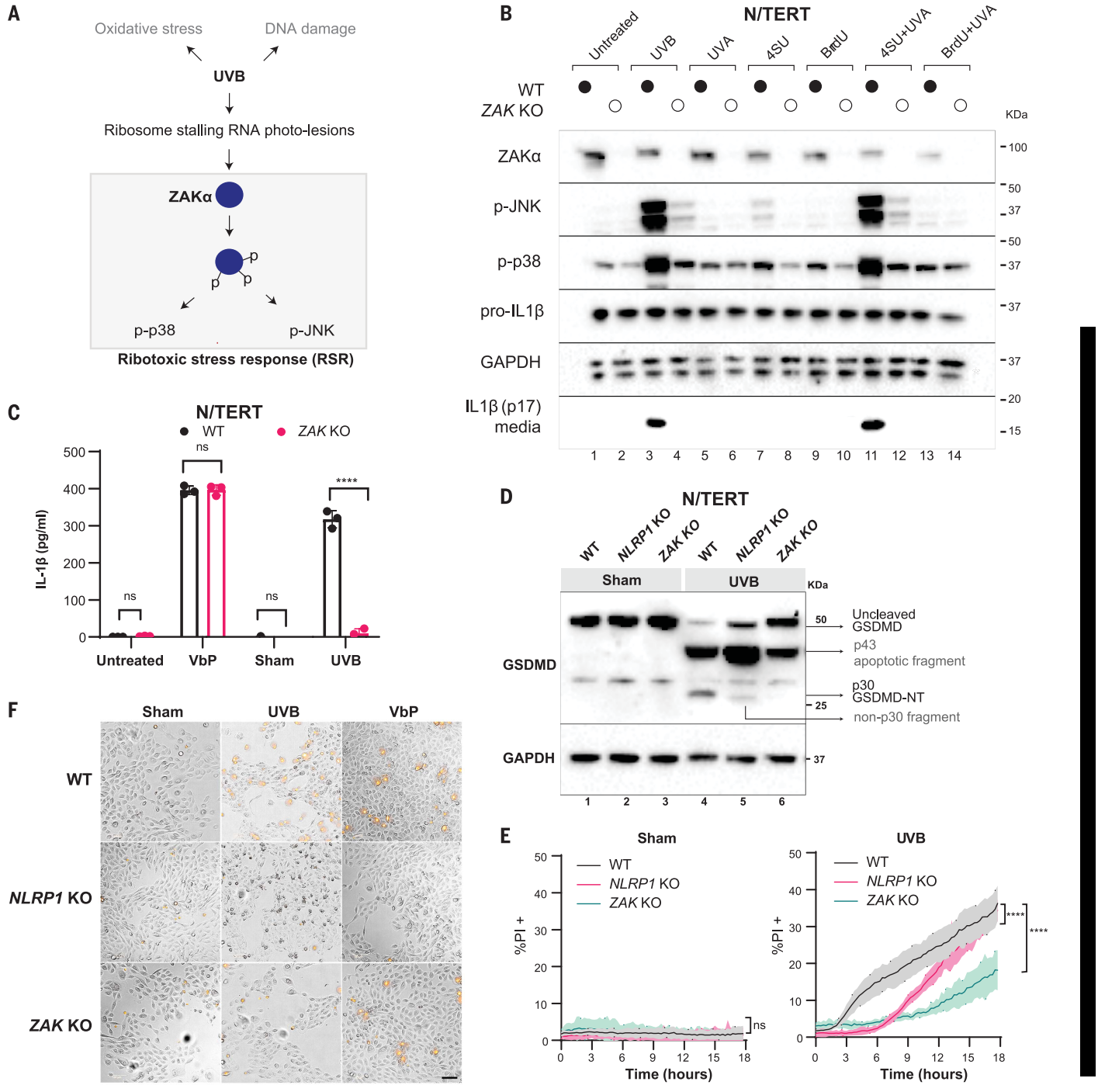


Fig. 1. ZAKα is required for UVB-triggered NLRP1 inflammasome activation. (**A**) Schematic indicating the types of cellular damage caused by UVB irradiation. UVB activates RSR signaling through ZAKα. (**B**) Immunoblot of WT N/TERT cells or *ZAK KO* (sg4) N/TERT cells treated with the indicated combinations of photosensitizer (10 μM) for 4 hours with or without UVA. UVB (100 mJ/cm^2) was used as a positive control. Asterisk indicates residual signal after membrane stripping. (**C**) IL-1β enzyme-linked immunosorbent assay (ELISA) of WT or *ZAK KO* (sg4) N/TERT cells after VbP (3 μM) treatment, sham irradiation, or UVB (100 mJ/cm^2) irradiation. Cell culture media were collected 24 hours later. (**D**) GSDMD immunoblot of WT, *NLRP1* KO, and *ZAK KO* (sg4) N/TERT cells treated with UVB (100 mJ/cm^2) or sham irradiated. Cell lysates were harvested 24 hours later. Different GSDMD cleavage fragments are shown by black arrows. Note that the GSDMD antibody used in this experiment recognizes all GSDMD-cleaved products. In *NLRP1* KO cells, UVB leads to a weak band <30 kDa. (**E**) Quantification of the percentage of PI-positive WT, *NLRP1* KO, and *ZAK KO* (sg4) N/TERT cells after sham irradiation or UVB (100 mJ/cm^2) irradiation. (**F**) Representative images of PI inclusion 5 hours after irradiation from three independent experiments. Scale bar, 100 μm. Error bars represent SEMs from three biological replicates, where one replicate refers to an independent seeding and treatment of the cells. Statistical significance was calculated by two-way ANOVA followed by Sidak's test for multiple pairwise comparisons in (C) and two-tailed Kolmogorov–Smirnov test at 95% confidence interval in (E). ns, nonsignificant; ****P < 0.0001.

full-length ZAKα induced a marked increase in ASC-GFP specks when coexpressed with NLRP1. By contrast, neither ZAKβ nor any of the ZAKα mutants defective in kinase function or sensing of ribosome stalling or collisions exhibited this function (fig. S5, C and D). Thus, the ribosome binding and RSR sensing domains of ZAKα are required for both RSR and NLRP1 inflammasome activation.

We next tested the effects of established ZAKα-activating toxins anisomycin (ANS) and doxyvinenol (DON) (table S1) (*33*). Additionally, we predicted that hygromycin (HYGRO), the effect on ZAKα was unknown at the time of our study, would also function as a ZAKα-RSR activator. To control for protein synthesis inhibition, we tested several translational inhibitors blasticidin (BLAST), puromycin (PURO), emetine, and G418, which target different sites of the ribosome and do not activate ZAKα. A nonspecific cytotoxic drug, staurosporine (STS),

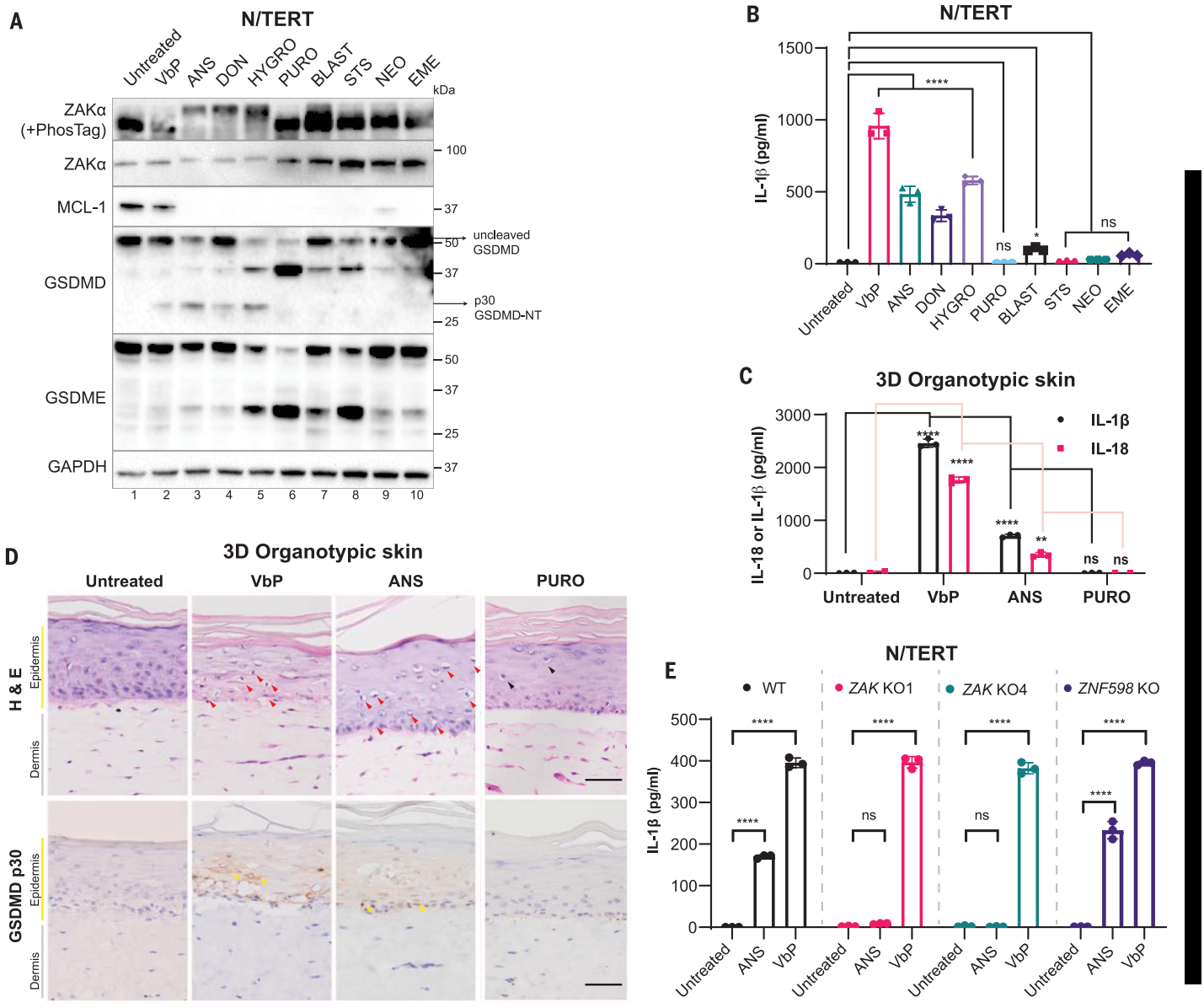


Fig. 2. ZAKα-activating compounds induce NLRP1-driven pyroptosis. (**A**) Immunoblot of N/TERT cell lysate after treatment with the indicated drugs. ZAKα phosphorylation was detected after 3 hours of drug treatment, and immunoblots for MCL-1, GSDMD, and gasdermin E (GSDME) were performed using samples 24 hours afte treatment. (**B**) IL-1β ELISA of N/TERT cell media 24 hours after treatment with the indicated drugs at concentrations specified in (A). Note that a smaller volume of medium and a higher number of cells were used in this experiment, accounting for the overall higher concentration of IL-1β. (**C**) IL-1β and IL-18 ELISA of growth media collected from three-dimensional (3D) organotypic skin cultures treated with the indicated drugs. (**D**) Hematoxylin and eosin (H&E) and cleaved GSDMD-NT (p30 specific) immunostaining of 3D organotypic skin cultures treated with the indicated drugs in (D). Scale bar, 100 μm. Red arrows indicate keratinocytes with diminished eosin but dense hematoxylin staining that were abundant in VbP- and ANS-treated cultures. Black arrows indicate putatively apoptotic cells with low H&E staining that were abundant in PURO-treated samples. Yellow arrows indicate membranous GSDMD p30 staining. Images represent one of three independent organotypic skin cultures. (**E**) IL-1β ELISA of culture media from N/TERT cells of the indicated genotypes after 24 hours of drug treatment. VbP was used at 3 μM and ANS at 1 μM. Error bars represent SEMs from three biological replicate experiments, where one replicate refers to an independent seeding and treatment of the cells. Statistical significance was calculated by two-way ANOVA followed by Dunnett's test for multiple pairwise comparisons in (B) and (C) and by Sidak's test in (E). *P < 0.05, **P < 0.01, ****P < 0.0001.

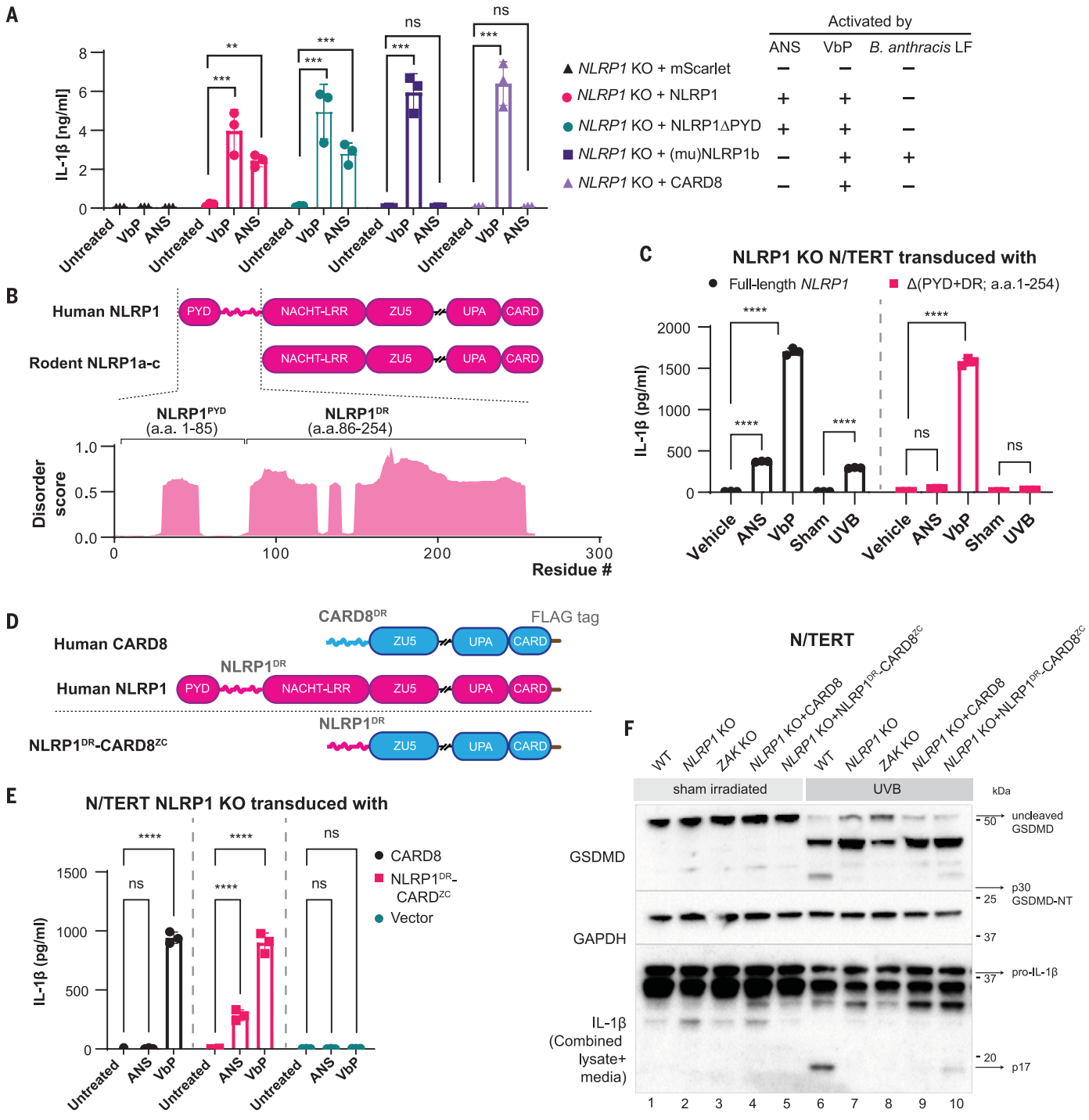


Fig. 3. A disordered linker region of human NLRP1 selectively mediates ZAKα-dependent activation. (**A**) IL-1β ELISA from *NLRP1* KO N/TERT cells reconstituted with the indicated NLRP1 variants and CARD8 treated with ANS (1 μM) and VbP (3 μM). Note that this experiment was performed independently using higher cell numbers. (**B**) Comparison between the domain structures of human NLRP1 and rodent NLRP1a-c. The predicted disorder score was calculated for amino acids 1 to 300 of human NLRP1. (**C**) IL-1β ELISA from *NLRP1*-KO N/TERT cells reconstituted with GFP-full-length NLRP1 or NLRP1 lacking PYD+DR (amino acids 1 to 254). Cells were treated with the indicated drugs or sham-irradiated or UVB-irradiated and harvested 24 hours after treatment. (**D**) Comparison of the domain arrangements of human NLRP1 and CARD8 and the engineered hybrid sensor referred to as NLRP1^{DR}-CARD8^{ZC}. (**E**) IL-1β ELISA from *NLRP1*-KO N/TERT cells transduced with CARD8 or NLRP1^{DR}-CARD8^{ZC} and treated with 1 μM ANS or 3 μM VbP for 24 hours. (**F**) GSDMD and IL-1β immunoblot from the cells in (E), along with WT or *ZAK* KO N/TERT cells irradiated with UVB. The GSDMD antibody recognizes both full-length and cleaved forms, including p43 and p30. The IL-1β immunoblot was performed with samples that combined lysate and 10 times concentrated medium. The error bars represent SEMs from three biological replicates, where one replicate refers to an independent seeding and treatment of the cells. Statistical significance in (A), (C), and (E) was calculated by two-way ANOVA followed by Sidak's test for multiple pairwise comparisons. $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

was used to exclude the possibility that NLRP1 was inadvertently activated by apoptosis. ANS, DON, and HYGRO strongly induced ZAKα phosphorylation in N/TERT cells (Fig. 2A). Among the other cytotoxic drugs, only BLAST induced moderate ZAKα phosphorylation. By contrast, the level of antiapoptotic protein MCL-1, which is a sensor for translational inhibition (*34*), was reduced by all drugs tested (Fig. 2A). There was a correlation between the level of ZAKα phosphorylation and inflammasome activation, with ANS, HYGRO, and DON acting as strong inducers of IL-1β secretion and GSDMD p30 cleavage in N/TERT cells and BLAST as a weak inducer (Fig. 2, A and B, and fig. S5, E and F). None of the non-ZAKα-activating drugs induced detectable IL-1β p17 secretion or GSDMD p30 cleavage despite significant cell death (Fig. 2, A and B). ANS also induced IL-1β and IL-18 secretion and GSDMD $p30^+$ cells with condensed nuclei and strong eosin staining in organotypic human skin cultures (Fig. 2, C and D). Thus, ZAKα-activating compounds, such as ANS, DON, and HYGRO are NLRP1 inflammasome activators. Moreover, this property is not caused by a general inhibition of translation.

Additionally, knocking out ZAKα, NLRP1, or components of the canonical inflammasome (but not NLRP3 or the ribosome-associated protein quality control sensor ZNF598) abrogated ANS-dependent pyroptosis in N/TERT cells (Fig. 2E and figs. S5G and S6A). ANS-dependent NLRP1 activation was also sensitive to the NEDD8/cullin inhibitor MLN4924 (fig. S6B). ANS also induced NLRP1 CT oligomerization and the formation of ASC-GFP specks in 293T-ASC-GFP-NLRP1 reporter cells, two established readouts for inflammasome activation, without affecting DPP9 protease activity (fig. S7, A to D). ANS could also induce inflammasome-driven pyroptosis in foreskin keratinocytes, bronchial epithelial cells, and aortic endothelial cells (fig. S8, A to E), but not in MV-4-11 cells, which use CARD8 as the primary inflammasome sensor (fig. S8F). Thus, ZAKα-activating ribotoxins, exemplified by ANS, DON, and HYGRO, function as bona fide NLRP1 activators.

We next reconstituted *NLRP1* KO N/TERT cells with human NLRP1, NLRP1ΔPYD, murine NLRP1B (muNLRP1B), or human CARD8 (Fig. 3A and fig. S9, A and B). When overexpressed, all of these heterologous sensors restored VbP-induced IL-1β secretion (Fig. 3A), as reported previously (*13*). By contrast, ANS only induced IL-1β in cells rescued with human NLRP1 or NLRP1ΔPYD (Fig. 3A) but not murine NLRP1 or human CARD8. Thus, ZAKα-activating molecules such as ANS are specific triggers for human NLRP1. In further support for its species specificity, ANS did not induce IL-1β release in murine bone marrow–derived macrophages in a muNLRP1-dependent manner (fig. S9, C and D) (*35*, *36*).

Human NLRP1 harbors a unique N-terminal extension encompassing the nonsignaling PYD and an extended linker, which is absent in rodent NLRP1 orthologs and CARD8. This linker region is predicted to be intrinsically disordered (Fig. 3B) (*37*). Recent work documented a critical role of a similarly disordered linker region in CARD8 inflammasome activation (*28*). We therefore tested the role of the NLRP1 linker region (amino acids 86 to 254, hereafter referred to as $NLRP1^{DR}$). The deletion of the $NLRP1^{DR}$, but not NLRP1 PYD, abrogated UVB- and ANS-triggered pyroptosis (Fig. 3, A and C, and fig. S10, A and B). By contrast, VbP-dependent NLRP1 inflammasome activation was unaffected. Similar results were obtained in 293T-ASC-GFP cells (fig. S10, C and D). In addition, $NLRP1^{DR}$ deletion attenuated NT degradation after UVB or ANS treatment (fig. S10B). Although the deletion of $NLRP1^{DR}$ also affected VbP-induced NT degradation, $NLRP1^{DR}$ deletion did not affect VbP-triggered pyroptosis (Fig. 3C and fig. S10B), suggesting that VbP-induced NLRP1 activation involves domains other than $NLRP1^{DR}$. Thus, $NLRP1^{DR}$ is selectively required for ZAKα-dependent NLRP1 inflammasome activation.

To further dissect the function of $NLRP1^{DR}$, we constructed a hybrid human inflammasome sensor (called "$NLRP1^{DR}$-$CARD8^{ZC}$" with a C-terminal FLAG tag) by fusing $NLRP1^{DR}$ to the signaling domains of CARD8 (ZU5-UPA-CARD) (Fig. 3D). Because CARD8 itself is insensitive to UVB and ANS (Fig. 3A), any neomorphic gain in inflammasome response to ANS and UVB could be attributed to $NLRP1^{DR}$. When $NLRP1^{DR}$-$CARD8^{ZC}$ was expressed in *NLRP1* KO N/TERTs, inflammasome activation in response to UVB (Fig. 3, E and F, and fig. S11, A and B), ANS (figs. S11C and S12, A and B), and HYGRO (fig. S11C) was restored. Thus, the disordered linker region is a necessary and sufficient determinant for NLRP1 to sense ZAKα-activating agents. $NLRP1^{DR}$ as a GFP fusion protein ($NLRP1^{DR}$-GFP) showed increased fluorescence 24 hours after ANS treatment for unknown reasons (fig. S13A), suggesting that $NLRP1^{DR}$ does not undergo proteasomal degradation and that other unknown structural elements on NLRP1 must be necessary for NT degradation.

We observed a marked band shift for $NLRP1^{DR}$-GFP by immunoblot whenever the cells were treated with UVB or ANS (Fig. 4A). This band shift was sensitive to post-lysis treatment with lambda phosphatase (fig. S14A) and could be eliminated by mutating all the serine and threonine residues to alanine (a mutant hereafter referred to as "Stless") (fig. S13B), which confirmed that it was caused by phosphorylated $NLRP1^{DR}$. PhosTag SDS-PAGE further revealed that $NLRP1^{DR}$ was significantly phosphorylated in unstimulated cells and became further phosphorylated by ANS and UVB (Fig. 4A). We hereafter refer to ANS- or UVB-dependent $NLRP1^{DR}$ phosphorylation as hyperphosphorylation. Among all the drugs tested, only ZAKα-activating compounds could induce $NLRP1^{DR}$ hyperphosphorylation (fig. S13C). In *ZAK*-KO N/TERT cells, ANS- and UVB-induced $NLRP1^{DR}$ hyperphosphorylation was abrogated (Fig. 4A). ZAKα-driven NLRP1 hyperphosphorylation was specific, because ASC was not phosphorylated by ANS or VbP treatment (fig. S13D). In orthogonal experiments, coexpression of wild-type (WT) ZAKα induced a large band shift of full-length NLRP1 in 293T cells on PhosTag gel, which was diminished by the deletion of $NLRP1^{DR}$ or the ZAKα kinase–dead mutation (p.K45A) (fig. S14, B and C). Thus, ZAKα, when activated either by overexpression or ribotoxic stress, hyperphosphorylates NLRP1 within the $NLRP1^{DR}$.

Mutating the serine/threonine residues to alanine within a short stretch of $NLRP1^{DR}$ (aminio acids 121 to 196) abrogated NLRP1 activation by UVB but did not affect VbP-driven IL-1β secretion (fig. S16). This suggested that this region harbors critical ZAKα-dependent phosphorylation sites. Recombinant ZAKα was sufficient to phosphorylate SNAP-tagged $NLRP1^{DR}$ purified from bacteria (Fig. 4B), indicating that NLRP1 is a direct substrate of ZAKα. Mass spectrometry of excised p-$NLRP1^{DR}$ bands after coincubation with ZAKα identified seven distinct ZAKα phosphorylation sites. These sites are clustered in two identical motifs of the sequence PTSTAVL (Fig. 4B and data S1), which does not exist in any other protein in the human proteome annotated in Swiss-Prot. Therefore, we propose naming this sequence the ZAKα motif. The second ZAKα motif (motif #2) falls within the crucial region (amino acids 121 to 196) identified by alanine scanning (fig. S16) and can be phosphorylated by ZAKα in vitro in the absence of any other phosphorylation sites (fig. S18A). Mutating the three serine/threonine residues in this motif (amino acids T178A, S179A, T180A, resulting in the NLRP1 3A mutant) eliminated UVB- and ANS-induced pyroptosis in reconstituted *NLRP1*-KO N/TERT cells, but had no effect on VbP-dependent pyroptosis (Fig. 4, D to F, and fig. S17, A and B). Thus, a single phosphorylation site in the ZAKα motif within $NLRP1^{DR}$ controls ZAKα-driven NLRP1 activation.

Given that ZAKα activates multiple SAPKs, we tested whether other kinases also contributed to ZAKα-induced NLRP1 activation. In agreement with previous results (*38*), multiple p38 inhibitors blocked ANS- and UVB-induced IL-1β secretion (fig. S17C) without altering VbP-dependent NLRP1 activation. By contrast, inhibitors of JNK, TAK1, SYK, and MK2 did not affect ANS-driven NLRP1 activation (fig. S17, C and D). Both p38α and p38β could phosphorylate recombinant $NLRP1^{DR}$, including residues within the ZAKα motifs (Fig. 4B; figs. S17F and S18, A and B; and data S1). However,

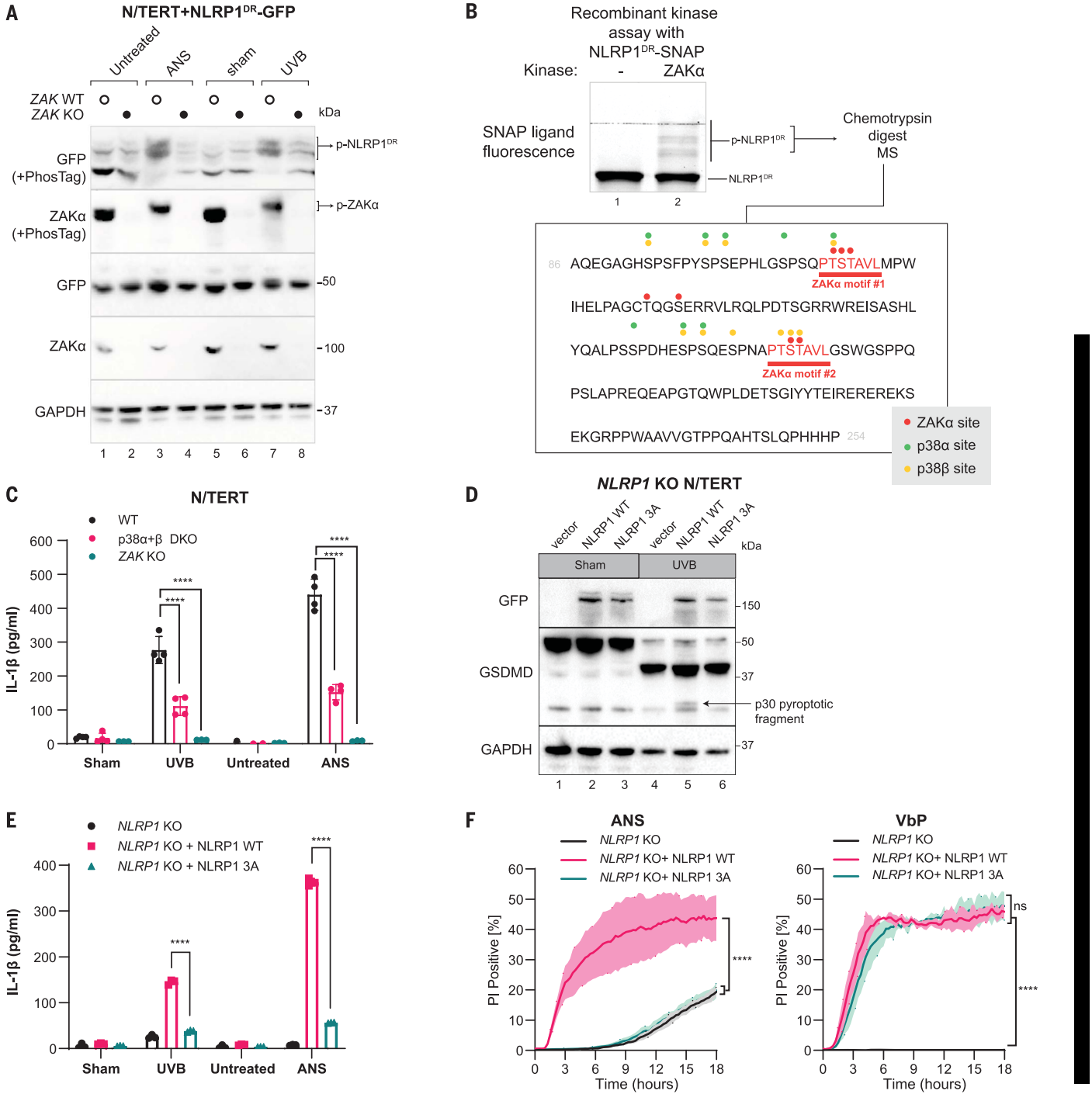


Fig. 4. Hyperphosphorylation of NLRP1^{DR} by ZAKα and p38 activates NLRP1. (**A**) Immunoblot after SDS-PAGE or PhosTag SDS-PAGE of WT or *ZAK*-KO N/TERT cells expressing NLRP1^{DR}-GFP. Cells were harvested 2 hours after ANS treatment or UVB irradiation. (**B**) Recombinant SNAP-tagged NLRP1^{DR} was incubated with recombinant ZAKα in a standard kinase reaction for 30 min. NLRP1^{DR} phosphorylation was visualized with SNAP ligand fluorescence (TMR) on a PhosTag-containing SDS-PAGE gel. (**C**) IL-1β ELISA from WT, *MAPK14 MAPK11* DKO (denoted as p38α+β DKO) and *ZAK* KO N/TERT cells 24 hours after UVB irradiation or ANS treatment. (**D**) GFP and GSDMD immunoblot of *NLRP1*-KO N/TERT cells expressing full-length WT NLRP1 or full-length NLRP1 T178A, S179A, or T180A (3A) mutants after UVB irradiation. All constructs were fused with GFP at the N terminus. GSDMD p30 is marked with a black arrow. Cells were harvested 24 hours after UVB irradiation. (**E**) IL-1β ELISA from *NLRP1*-KO N/TERT cells expressing full-length WT NLRP1 or full-length NLRP1 T178A, S179A, or T180A (3A) mutants 24 hours after UVB irradiation or ANS treatment. (**F**) Quantification of the percentage of PI^{+}*NLRP1*-KO N/TERT cells expressing full-length WT NLRP1 or full-length NLRP1 T178A, S179A, or T180A (3A) mutants in the presence of ANS or VbP. Images were acquired at 15-min intervals for 18 hours. Error bars represent SEMs from three biological replicates, where one replicate refers to an independent seeding and treatment of the cells. Statistical significance was calculated by two-way ANOVA followed by Sidak's test for multiple pairwise comparisons in (C) and (E) and two-tailed Kolmogorov–Smirnov test at 95% confidence interval in (F). ****P < 0.0001.

p38 inhibitors showed only a modest effect on ANS-induced NLRP1^{DR} hyperphosphorylation (fig. S17E), which was completely abrogated by ZAKα inhibitors (fig. S17E). A double KO of p38α and p38β (p38α+β dKO) in N/TERTs produced a significant but incomplete reduction of IL-1β secretion after UVB and ANS treatment (Fig. 4C and fig. S18, C and D). Thus, p38 kinases contribute to NLRP1 inflammasome activation induced by RSR-inducing agents, but their roles are less critical than that of ZAKα. Because p38α and p38β are strictly downstream of ZAKα in RSR signaling, we hypothesize that ZAKα is the initiating kinase that phosphorylates NLRP1, whereas p38 kinases subsequently reinforce or amplify this response.

UVB irradiation leads to caspase-1–dependent pyroptotic cell death and IL-1β secretion in human skin keratinocytes (*15*, *27*, *39*). Initially thought to be an NLRP3-associated phenomenon, recent work has provided convincing evidence that NLRP1, rather than NLRP3, is the primary inflammasome sensor for UVB. In this work, we have identified the key events controlling UVB-triggered NLRP1 inflammasome activation. By inducing cellular RNA photolesions that stall ribosomes, UVB activates the RSR kinase ZAKα, which, together with its downstream effector p38, phosphorylates the human-specific disordered linker region of NLRP1. A single phosphorylation site within the ZAKα motif identified here is sufficient to control NLRP1 activation in an DPP8/9-independent manner. Thus, human NLRP1 is a very versatile sensor protein that can integrate multiple signals through its discrete domains. We have expanded the repertoire of known human NLRP1 agonists to include multiple microbial ribotoxins such as ANS and DON. The same results have been independently reported by a concurrent study (*40*). Our work raises several questions to be addressed by future studies. For instance, it is currently unclear how ZAKα-driven hyperphosphorylation activates NLRP1 by accelerating the functional degradation of NLRP1 NT. Our findings also suggest that RSR signaling may play a role in antimicrobial defense and IL-1–driven inflammation in vivo. If proven, then pharmacologically targeting the ZAKα-NLRP1 axis may prove beneficial in treating human inflammatory disorders.

REFERENCES AND NOTES

1. V. A. K. Rathinam, K. A. Fitzgerald, *Cell* **165**, 792–800 (2016).
2. P. Broz, V. M. Dixit, *Nat. Rev. Immunol.* **16**, 407–420 (2016).
3. S. M. Man, T.-D. Kanneganti, *Immunol. Rev.* **265**, 6–21 (2015).
4. C. Y. Taabazuing, A. R. Griswold, D. A. Bachovchin, *Immunol. Rev.* **297**, 13–25 (2020).
5. D. A. Bachovchin, *Mol. Cell* **81**, 423–425 (2021).
6. P. S. Mitchell, A. Sandstrom, R. E. Vance, *Curr. Opin. Immunol.* **60**, 37–45 (2019).
7. F. L. Zhong *et al.*, *J. Biol. Chem.* **293**, 18864–18878 (2018).
8. M. C. Okondo *et al.*, *Cell Chem. Biol.* **25**, 262–267.e5 (2018).
9. D. C. Johnson *et al.*, *Nat. Med.* **24**, 1151–1156 (2018).
10. L. R. Hollingsworth *et al.*, *Nature* **592**, 778–783 (2021).
11. M. Huang *et al.*, *Nature* **592**, 773–777 (2021).
12. K. S. Robinson *et al.*, *Science* **370**, eaay2002 (2020).
13. S. Bauernfried, M. J. Scherr, A. Pichlmair, K. E. Duderstadt, V. Hornung, *Science* **371**, eabd0811 (2021).
14. B. V. Tsu *et al.*, *eLife* **10**, e60609 (2021).
15. G. Fenini *et al.*, *J. Invest. Dermatol.* **138**, 2644–2652 (2018).
16. K. M. Cirelli *et al.*, *PLOS Pathog.* **10**, e1003927 (2014).
17. J. Chavarría-Smith, R. E. Vance, *PLOS Pathog.* **9**, e1003452 (2013).
18. S. E. Ewald, J. Chavarria-Smith, J. C. Boothroyd, *Infect. Immun.* **82**, 460–468 (2014).
19. A. Sandstrom *et al.*, *Science* **364**, eaau1330 (2019).
20. A. J. Chui *et al.*, *Science* **364**, 82–85 (2019).
21. H. Xu *et al.*, *EMBO J.* **38**, e101996 (2019).
22. F. L. Zhong *et al.*, *Cell* **167**, 187–202.e17 (2016).
23. J. Sand *et al.*, *Cell Death Dis.* **9**, 24 (2018).
24. S. B. Drutman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19055–19063 (2019).
25. S. Grandemange *et al.*, *Ann. Rheum. Dis.* **76**, 1191–1198 (2017).
26. C. B. Levandowski *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2952–2956 (2013).
27. G. Sollberger *et al.*, *J. Invest. Dermatol.* **135**, 1395–1404 (2015).
28. A. J. Chui *et al.*, *Cell Rep.* **33**, 108264 (2020).
29. P. Karran, R. Brem, *DNA Repair* **44**, 178–185 (2016).
30. A. C. Vind *et al.*, *Mol. Cell* **78**, 700–713.e7 (2020).
31. C. C.-C. Wu, A. Peterson, B. Zinshteyn, S. Regot, R. Green, *Cell* **182**, 404–416.e14 (2020).
32. D. Markowitz *et al.*, *Mol. Cancer Ther.* **15**, 1799–1808 (2016).
33. A. C. Vind, A. V. Genzor, S. Bekker-Jensen, *Nucleic Acids Res.* **48**, 10648–10661 (2020).
34. M. H. Orzalli *et al.*, *Immunity* **54**, 1447–1462.e5 (2021).
35. M. L. Vyleta, J. Wong, B. E. Magun, *PLOS ONE* **7**, e36044 (2012).
36. B. Briard *et al.*, *Nature* **588**, 688–692 (2020).
37. J. Hanson, K. K. Paliwal, T. Litfin, Y. Zhou, *Genomics Proteomics Bioinformatics* **17**, 645–656 (2019).
38. G. Fenini *et al.*, *J. Invest. Dermatol.* **138**, 1380–1390 (2018).
39. L. Feldmeyer *et al.*, *Curr. Biol.* **17**, 1140–1145 (2007).
40. L.-M. Jenster *et al.*, bioRxiv 477423 [Preprint] (2022); https://doi.org/10.1101/2022.01.24.477423.

ACKNOWLEDGMENTS

We thank the SRIS Asian Skin Biobank (ASB), especially A. Alimat, J. Lee, and A. Yap, for generating and providing human primary skin cells; the staff at the NTU Optical Bio-imaging Centre and the NTU SBS Mass Spec Core Facility, especially B. Kannan and W. Meng, for technical assistance with multiple experiments; T. Nguyen for contributions in multiple experiments; and especially B. Reversade, K. Lay, and S. Xue for detailed critiques of this manuscript. K.S.R. would like to dedicate her efforts to Colin Brian Gibson, who sadly lost his fight to melanoma in May 2021.

Funding: F.L.Z.'s laboratory is funded by the National Research Foundation, Singapore (NRF-NRFF11-2019-0006) and a Nanyang Assistant Professorship (NAP). K.W.C.'s laboratory is funded by National University of Health Sciences (NUHS) internal grant funding (NUHSRO/2020/117/STARTUP/04). L.H.'s laboratory is funded by the National Research Foundation, Singapore (NRF-NRFF2017-05), and the Howard Hughes Medical Institute (HHMI IRSP 5500873P). K.T. and J.C. are funded by the Industry Alignment Fund Pre-Positioning Programme (IAF-PP) H18/01a0/016 "Asian Skin Microbiome Program." A.S.B. is funded by A*STAR through a core fund and under the IAF-PP project H1701a0004 (phase 2: SRIS@Novena). V.H.'s laboratory is funded by ERC-2020-ADG ENGINES (101018672). R.S. is funded by an A*STAR core fund. B.W. and F.L.Z. are funded by a National Medical Research Council (NMRC)–Open Fund – Individual Research Grant (OFIRG) from the Ministry of Health, Singapore (MOH-000382). **Author contributions:** F.L.Z. conceived of and supervised this study. K.S.R. and G.A.T designed and performed experiments, analyzed the data, and co-wrote the manuscript with significant contributions from P.R., S.B., Z.S., M.J.F., Z.S.P., and S.B. All co-authors contributed to writing the manuscript. S.B. performed the N/TERT constitution experiments and analyzed the data with supervision from V.H. R.C. performed the recombinant protein purification and mass spectrometry identifying phosphorylation sites with supervision from B.W. C.R.H. performed the bone marrow–derived macrophage experiments with supervision from S.L.M. K.W.C. performed independent validation experiments with bone marrow–derived macrophages. R.N. and W.C. performed all endothelial cell experiments with supervision from L.H. C.K.L. performed cytokine analysis with supervision from J.E.C. C.B. and the Asian Skin Bank performed primary cell derivation and culture. K.Y.T. performed mass spectrometry experiments with supervision from R.S. **Competing interests:** S.L.M. is a scientific adviser for Odyssey Therapeutics and NRG Therapeutics. J.E.C. is a member of the board of directors for 3T Biosciences and is Chief Scientific Officer of Parker Institute for Cancer Immunotherapy. F.L.Z., K.S.R., G.A.T., and Z.S. are coinventors of a patent (PCT/SG2022/050086) based on this work. The remaining authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. N/TERT-1 cells were provided under a materials transfer agreement from the Brigham and Women's Hospital to the Institute of Medical Biology, A*STAR.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl6324
Materials and Methods
Figs. S1 to S18
Tables S1 and S2
Data S1
References
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 27 July 2021; accepted 14 June 2022
10.1126/science.abl6324

MOLECULAR SEPARATION

Discrimination of xylene isomers in a stacked coordination polymer

Liangying Li[1,2]†, Lidong Guo[1]†, David H. Olson[2], Shikai Xian[2,3], Zhiguo Zhang[1,4], Qiwei Yang[1,4], Kaiyi Wu[1], Yiwen Yang[1,4], Zongbi Bao[1,4]*, Qilong Ren[1,4], Jing Li[2,3]*

The separation and purification of xylene isomers is an industrially important but challenging process. Developing highly efficient adsorbents is crucial for the implementation of simulated moving bed technology for industrial separation of these isomers. Herein, we report a stacked one-dimensional coordination polymer {[Mn(dhbq)(H_2O)$_2$], H_2dhbq = 2,5-dihydroxy-1,4-benzoquinone} that exhibits an ideal molecular recognition and sieving of xylene isomers. Its distinct temperature-adsorbate–dependent adsorption behavior enables full separation of *p*-, *m*-, and *o*-xylene isomers in both vapor and liquid phases. The delicate stimuli-responsive swelling of the structure imparts this porous material with exceptionally high flexibility and stability, well-balanced adsorption capacity, high selectivity, and fast kinetics at conditions mimicking industrial settings. This study may offer an alternative approach for energy-efficient and adsorption-based industrial xylene separation and purification processes.

Xylene isomers are widely used as raw chemicals for manufacturing large-scale industrial commodities (*1*, *2*). Separation of xylene isomers through distillation is an energy-intensive process because of their nearly-the-same boiling points (table S1) (*3*). The dominant industrial method for large-scale separation of xylene isomers is based on simulated moving bed (SMB) using FAU-type zeolites as the stationary phase (*4*), which is favorably run at relatively high temperatures (393 to 523 K or 120° to 250°C) to ensure a sufficient mass-transfer rate and to reduce viscosity of the mobile phase as well as pressure drop across adsorbent beds connected in a series (*4*, *5*). However, these zeolite adsorbents suffer from low selectivity and/or uptake capacity as a result of limited porosity, structural tunability, and high rigidity (*6*). Therefore, developing stable and high-performance adsorbents represents an important current interest for the efficient separation of xylene isomers.

Solid adsorbents, such as MFI-type zeolites (*7*), molecular crystals (*8*–*10*), metal complexes (*11*, *12*), organic cages (*13*), and metal-organic frameworks (MOFs) (*14*–*20*), have been investigated for the separation of xylene isomers. A particular feature of dynamic MOFs is their framework flexibility, which has led to interesting and unexpected adsorption properties (*21*–*24*). Such structural flexibility, including breathing and gate-opening (Fig. 1A), has been utilized for the selective adsorption of xylenes (*25*–*28*); however, a complete separation of all three isomers by molecular recognition or sieving has not yet been realized.

We report a manganese-based stacked one-dimensional (1D) coordination polymer, Mn(dhbq)(H_2O)$_2$ (Mn-dhbq, H_2dhbq = 2,5-dihydroxy-1,4-benzoquinone), and its performance for discrimination and separation of xylene isomers in the temperature range of 303 to 393 K (30° to ~120°C). Mn-dhbq exhibits distinct changes in the interchain distances when exposed to xylene molecules, leading to a framework swelling (Fig. 1A) and discriminative adsorption behavior toward the three isomers.

Mn-dhbq was obtained as a brown-orange–colored microcrystalline solid (fig. S1), following the reported procedure with minor modifications (*29*, *30*). The structure of the as-synthesized Mn-dhbq was refined by Rietveld method using the room-temperature powder x-ray diffraction (PXRD) data (fig. S2), which are in good agreement with those reported by Kitagawa (table S3) (*29*, *30*). Each dhbq coordinates to two Mn^{2+} (Fig. 1B), and each Mn^{2+} bonds to four oxygens from two ligands and two oxygens from two coordinated water molecules, resulting in a strip-like 1D straight chain (Fig. 1C). The adjacent chains are linked by multiple hydrogen bonds, leading to a H-bonded 3D network with a hydrogen bond distance of 1.754 Å (Fig. 1, C and D). Upon activation, coordinated water molecules were removed, and the sample remained stable up to 748 K (475°C) (fig. S3), confirmed by the porosity analysis (fig. S4). The framework underwent a structure transformation accompanied by color change (fig. S5). This structural change is reversible (figs. S6 and S7). The Brunauer-Emmett-Teller (BET) surface area was estimated to be ~429 m^2/g with a pore size of ~5.6 Å (table S4). Based on Connolly surface and pore analysis by Material Studio 2017R2 (Accelrys Software, Inc.) using the as-made structure (with water molecules removed), the isolated tubular-shaped voids (Fig. 1, E and F) would be too small to hold xylene molecules. In reality, however, the activated Mn-dhbq shows strong adsorption toward xylenes, which suggests an expansion of pore space.

Single-component vapor adsorption isotherms of xylene isomers were collected at 303, 333, 363, and 393 K on Mn-dhbq powder samples. Molecular sieving–based full separation of xylene isomers was achieved with the affinity following the trend of para-xylene (PX) > meta-xylene (MX) > ortho-xylene (OX). At 333 K, the compound adsorbs both PX and MX but almost fully excludes OX (Fig. 2A). The uptake capacities are 185, 159, and 23 mg/g for PX, MX, and OX, respectively, at 1.05 kPa. This result reveals that OX can be sieved out at this temperature. At a higher temperature (393 K), both MX and OX are essentially excluded, with uptake capacities of 141, 17, and 22 mg/g, respectively, for PX, MX, and OX at 1.05 kPa (fig. S13). For effective separations, adsorption capacity is an equally important parameter as selectivity. Mn-dhbq takes up substantially higher amounts of PX than those of best-performing PX-selective adsorbents. For example, the maximum uptake amounts of MFI zeolite (131.4 mg/g, 298 K) (*7*), Zn-MOF (55 mg/g, 298 K) (*27*), and Cu-metallocycle (140 mg/g, 293 K) (*12*) are all considerably lower than that of Mn-dhbq (>200 mg/g) measured at a higher temperature (303 K). The temperature-dependent adsorption behavior of Mn-dhbq allows for separation of both PX/OX and PX/MX binary mixtures. Notably, the adsorption isotherms of PX exhibit steep slopes, reaching saturation at very low pressure, which suggests strong interactions of Mn-dhbq with PX. Distinct inflection points were observed in both MX and OX isotherms (figs. S10 to S13)—an indication of swelling of the Mn-dhbq structure in response to the external stimulus. The adsorption performance was retained for the formulated pellet samples (Fig. 2D and figs. S8, S9, and S14 to S16).

To assess the capability of Mn-dhbq for selective recognition of a specific xylene, we performed competitive adsorption measurements on xylene mixtures in both vapor and liquid phase. The data were analyzed by proton nuclear magnetic resonance (^{1}H-NMR) (Fig. 2B and figs. S17 to S28) and gas chromatography (figs. S29 to S32). For equimolar binary vapor mixtures, the relative adsorption selectivities of Mn-dhbq are 34.8/1 and 37.9/1 for PX/MX and PX/OX at 363 K, respectively (figs. S24 and S25), and 12.5/1 for MX/OX at 333 K (fig. S26). These values are higher than most previously reported numbers (Fig. 2E and table S2). Additionally, a high selectivity of 70.4/2.5/1 was obtained for an equimolar ternary vapor mixture of PX/MX/OX at 363 K (fig. S28). The

[1]Key Laboratory of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, P. R. China. [2]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA. [3]Hoffmann Institute of Advanced Materials, Shenzhen Polytechnic, Shenzhen 518055, P. R. China. [4]Institute of Zhejiang University-Quzhou, Quzhou 324000, P. R. China.
*Corresponding author. Email: baozb@zju.edu.cn (Z.B.); jingli@rutgers.edu (J.L.)
†These authors contributed equally to this work.

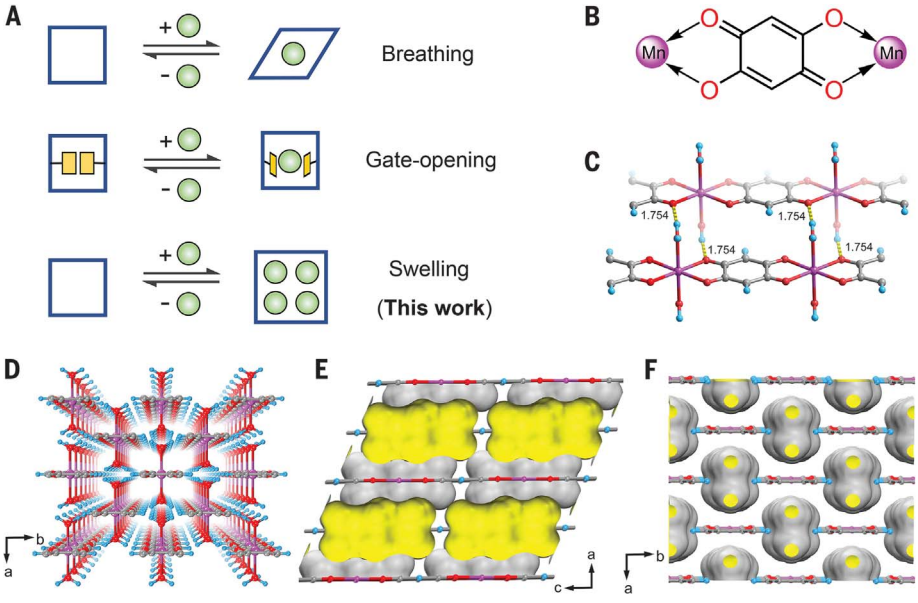


Fig. 1. The structure and pore properties of Mn-dhbq. (**A**) Representations of different modes of framework dynamics or flexibility: breathing, gate opening-closing or linker rotation, and swelling. (**B**) The coordination mode of the dhbq linker. (**C**) The hydrogen bonds between the two adjacent 1D chains within the framework (yellow dashed lines). (**D**) Perspective view of the crystal structure of as-synthesized Mn-dhbq along the *c* axis. (**E** and **F**) The pore spaces within the Mn-dhbq structure without the coordinated water molecules. The unit cell is kept the same as the as-synthesized structure. The pore sizes in both (E) and (F) are too small to allow xylene molecules to be adsorbed.

adsorbed amount of PX was slightly affected by the presence MX and OX, with the uptake decreased from 159 mg/g (single component) to 140 mg/g (ternary mixture) at 363 K.

Furthermore, we carried out static solid-liquid extraction experiments at temperatures between 303 and 393 K to evaluate separation performance in liquid phase. The results confirmed the distinct selectivities of Mn-dhbq in the order of PX >> MX >> OX, especially at the higher temperatures necessary to meet the mass-transfer rate requirement for industrial SMB processes. It is worth noting that although Cu-metallocycle demonstrates impressive selectivity at lower temperatures, the value continues to drop quickly as temperature increases (table S5 and figs. S33 to S36). At 363 and 393 K, its selectivity for ternary mixture PX/OX/MX (45.4/1.4/1 and 19.4/1.6/1) becomes substantially lower than that of Mn-dhbq (78.4/3.8/1 and 84.6/2.5/1). Notably, Mn-dhbq preferentially adsorbs PX from a quaternary liquid mixture (PX/OX/MX/EB = 22/22/50/6) at a composition identical to that used in a licensed UOP (Honeywell UOP, formerly known as UOP LLC or Universal Oil Products) Parex process (*31*). Its selectivity (171.1/4.9/1/18.4) is substantially higher than that of the Cu-metallocycle (86.6/3.3/1/6.6) at 393 K (figs. S31, S32, S35, and S36).

To evaluate the feasibility of Mn-dhbq for xylene separations under real-world conditions, breakthrough experiments were also conducted on pellet samples for vapor-phase binary and ternary xylene mixtures (figs. S37 to S40). The results are in excellent agreement with the trend determined from the pure-component vapor adsorption isotherms. At 363 K, PX can be fully separated from the PX/MX/OX ternary mixture, followed by a complete separation of MX from the binary mixture of MX/OX at 303 K (Fig. 2, C and F), which confirms that Mn-dhbq is the first adsorbent capable of full discrimination of the three xylene isomers in the vapor phase by molecular sieving mechanism.

High diffusion rates drastically decrease the adsorption-desorption time, which is crucial for industrial applications (*32*). In this regard, operation at elevated temperature is necessary to achieve fast kinetics and sufficient diffusion rate for practical separations of xylene mixtures. Adsorption kinetics experiments were performed to verify the mass-transfer rate of xylenes in Mn-dhbq using both gravimetric and volumetric methods, and time-dependent adsorption profiles of PX were measured between 303 and 393 K (figs. S41 to S46). The diffusion rate constant (D_c/r^2, per second) of PX in Mn-bhdq was obtained by fitting the kinetic curves according to Crank theory (*33*). The estimated values are 7.8×10^{-4} and 7.0×10^{-3} at 303 and 393 K, respectively (fig. S45), which outperform both BaX zeolite (4.8×10^{-3} at 453 K) (*34*) and Cu-metallocycle (7.6×10^{-5} and 4.5×10^{-4} at 303 and 393 K, respectively; fig. S46).

A detailed structural analysis of xylene-adsorbed Mn-dhbq was made using several techniques and methods. Although the samples remained crystalline, the broad PXRD peaks suggest high-level local disorders (figs. S47 to S50), making it difficult to solve their crystal structures. We also attempted to use the 3D microcrystal electron diffraction (micro-ED) method to reconstruct the structure, but without success (figs. S51 and S52). Nevertheless, because the 1D chains within the structure are not connected to each other by chemical bonds, but are only held together by weak interchain hydrogen bonds and/or van der Waals forces, the structure is highly flexible and can undergo a swelling to provide optimal constrained spaces for selective adsorption of xylene isomers (Fig. 3A). The large shifts in the (001), (100), and (020) diffraction planes of xylene-adsorbed Mn-dhbq from those of the as-synthesized and activated Mn-dhbq correlate to an increase in the interchain distance from 6.35 and 6.15 to 6.47 Å (Fig. 3C). The activated Mn-dhbq framework contains stacks of flexible 1D chains with a high concentration of open metal sites, offering favorable and size-specific adsorption binding sites between the chains.

Having the largest length (z) to width (x) ratio (1.38; table S1), PX fits very well in the space between the neighboring chains because of its similar geometry to the Mn-ligand-Mn segment, giving rise to strong affinity to Mn-dhbq (Fig. 3, B and E), consistent with the high uptake of PX of 208.4 mg/g at 303 K. OX, however, with the smallest z/x ratio (1.08; table S1), does not match well with the constrained space (Fig. 3, B and G) and thus has the least interaction with the chain. The interaction of MX with Mn-dhbq falls between that of PX and OX, in trend with its intermediate z/x ratio (1.23; table S1 and Fig. 3, B and F). The density functional theory (DFT) simulations provide a more quantitative assessment of the sorbate-sorbent interactions, illustrating that there are three different binding sites for xylene molecules. All three xylene molecules have comparable arene π-π stacking interactions (site I, green dotted lines; Fig. 3, E to G) with the aromatic ligand with similar distances (3.393, 3.317, and 3.346 Å for PX, MX, and OX, respectively), which suggests that this interaction may not be the dominant binding force. The distinct difference is found in the hydrogen bonds between the H atoms of the xylene methyl group and the O atoms from the ligand (site II, blue dotted lines; Fig. 3, E to G). There are four hydrogen bonds between PX and the 1D chain with distances in the range of 2.711 to 2.903 Å, whereas only three and two hydrogen bonds are formed between MX and OX and the chain because of the less-well-matched shape of the two molecules. The dipole-dipole interactions (site III,

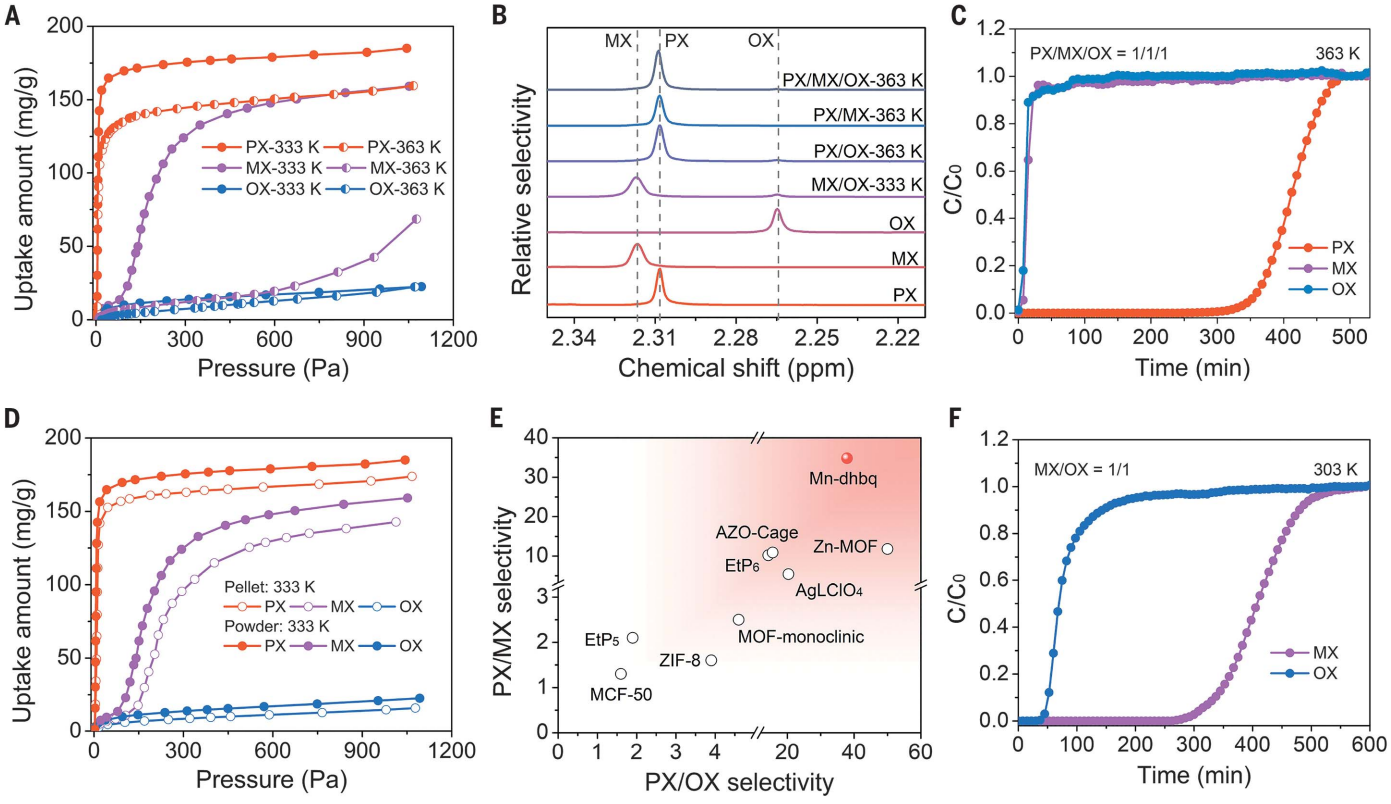


Fig. 2. Adsorption and separation of xylene isomers. (**A**) Single-component vapor adsorption isotherms of xylene isomers on Mn-dhbq at 333 and 363 K. (**B**) Magnified ^{1}H-NMR spectrum recorded using the $CDCl_3$ to extract xylene isomers from the decomposed Mn-dhbq by concentrated HCl solution that was priorly subjected to the equimolar binary or ternary vapor of xylene isomers at 333 and 363 K. ppm, parts per million. (**D**) The comparison of xylene vapor adsorption isotherms between Mn-dhbq powder and pellet samples at 333 K. (**E**) Comparison of adsorption selectivities for binary mixtures of PX/OX and PX/MX in vapor phase for Mn-dhbq and selected materials. For each reported material, only the highest and/or best values of selectivity were selected. (**C** and **F**) Breakthrough curves of an equimolar binary (F) and ternary (C) mixture of xylene isomers at 303 and 363 K for Mn-dhbq pellet samples. The partial pressure of xylenes in the feed gas used for binary and ternary breakthrough experiments is 2.4 and 3.9 kPa, respectively.

red dotted lines) between C_{methyl} and Mn^{2+} centers are also determined from molecular simulations (Fig. 3, E to G). For PX, there are two strong $C^{\delta-}\cdots Mn^{\delta+}$ dipole-dipole interactions between PX and the chain, with distances of 3.128 and 3.776 Å, respectively. For MX and OX, only one strong $C^{\delta-}\cdots Mn^{\delta+}$ dipole-dipole interaction exists (3.195 and 3.501 Å). The calculated binding energies of PX, MX, and OX (95.1, 92.8, and 87.1 kJ/mol) are fully consistent with the extent of three interactions and with the same descending order of the adsorption heats measured by differential scanning calorimeter (DSC) method (fig. S53). Additionally, thermogravimetric analysis of the xylene-loaded Mn-dhbq samples also confirmed the order of interactions (PX > MX > OX). The weakest-binding OX was lost from the sample at ~30°C, even though it has the highest boiling point among the three isomers. The weight loss of MX happened at ~95°C, followed by the loss of PX at ~126°C for MX- and PX-adsorbed Mn-dhbq (Fig. 3D and figs. S54 to S56). The relative binding strengths of the three xylene isomers correlate well with the observed phenomenon (*22*): PX is adsorbed at all experimental temperatures because its interaction with the Mn-dhbq framework is sufficiently strong to force the molecule to enter the constrained space at all temperatures. The intermediate binding energy between MX and Mn-dhbq is insufficient to allow the molecule to enter the voids at higher temperature (393 K). However, decreasing the temperature to a certain value (i.e., 363 K or below) will infuse sufficient energy to push MX into the adsorption sites. With the lowest binding energy, OX can only take up the sites at even lower temperatures (≤303 K).

Among reported sorbents for gas-vapor separations, many have poor chemical stability, especially those with open metal sites. By contrast, Mn-dhbq exhibits high resistance toward water, heat, and air over long periods of time, as confirmed by PXRD and porosity analysis (figs. S57 and S58). The porosity parameters of the activated Mn-dhbq sample remained nearly the same after being soaked in pure PX solution at 393 K for 24 hours or in boiled water for 1 week (fig. S58 and table S4). The in situ PXRD analysis on Mn-dhbq at various temperatures also points to its suitability for industrial xylene separations at relatively high temperatures, whereas Cu-metallocycle begins to show degradation at 433 K and loses crystallinity completely at 473 K (figs. S59 and S60). The PX uptake capacity (~147 mg/g) of Mn-dhbq remains unchanged after 20 consecutive adsorption-desorption cycles at 393 K (Fig. 4, A and B, and figs. S61 and S62). No obvious loss of PX uptake was observed after 10 consecutive adsorption cycles in the liquid phase at 393 K and subsequent desorption under nitrogen flow at 493 K (fig. S63). The Mn-dhbq structure remained intact with a stable cycling capacity of ~150 mg/g, whereas Cu-metallocycle lost its adsorption capacity and crystallinity after one cycle under identical conditions (figs. S63 and S64). The adsorbed PX was extracted from Mn-dhbq at 433 K (160°C, 1 hour), using 1,4-diethylbenzene (DEB) as desorbent, after adsorption of a quaternary liquid mixture of PX/OX/MX/EB (22/22/50/6) at 393 K (120°C). The purity of PX reached >97% with a recovery of 71% in a single desorption cycle (fig. S65). These results confirm the robustness and recyclability of Mn-dhbq. In

addition to the strong Mn-O bonds within the 1D chains, the flexible chain stacking structure that allows interchain rearrangement to quickly respond to the stimulus may be responsible for the reversible transformability between nonporous and porous states, leading to the high chemical stability of Mn-dhbq.

The commercially available raw chemicals and environmentally friendly method used to prepare Mn-dhbq are amenable for scaled-up synthesis. We achieved this by a simple one-pot reaction at room temperature using water as the only solvent. About 0.23 kg of Mn-dhbq was collected in a single batch (Fig. 4D) and formulated into pellets. The pellet sample exhibits high crystallinity (figs. S66 and S67) and similar separation performance as those of the powder samples (Fig. 2, C and F, and Fig. 4C).

A porous and robust coordination polymer, Mn-dhbq, having hydrogen-bonded 1D chains, multiple open-metal sites, rich π-electrons, and structural flexibility, can selectively recognize specific xylene isomers and is capable of full separation of OX, MX, and PX isomers through molecular sieving. Guest-induced structural rearrangement is important for the preferential interaction with specific xylene isomers. Two key properties of Mn-dhbq are the high separation selectivity for PX over MX and OX in both vapor and liquid phases, especially at higher temperatures, and the separation of all three xylene isomers in vapor phase through molecular sieving.

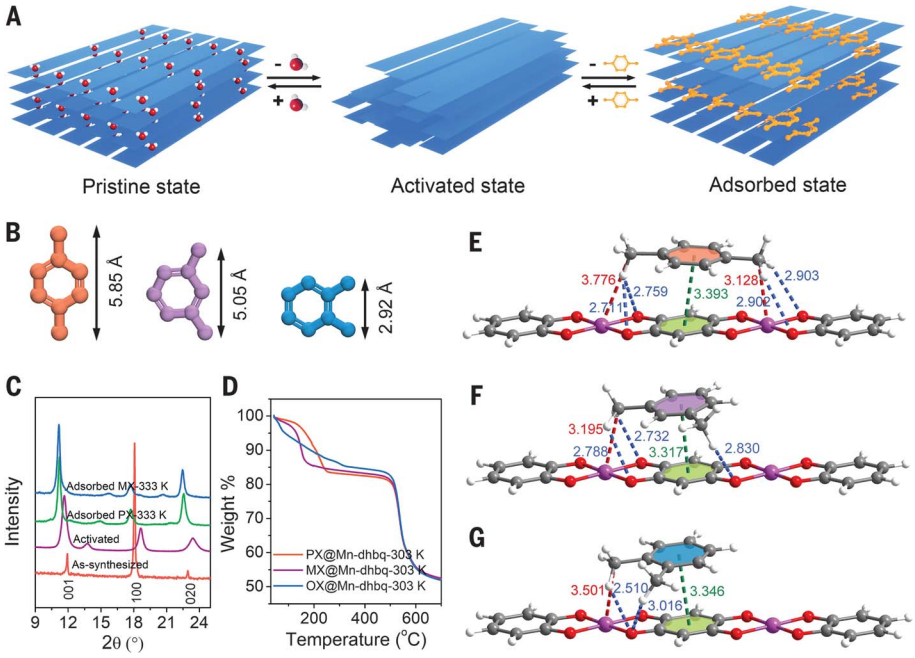


Fig. 3. Mechanism of framework swelling and xylene–Mn-dhbq interaction. (**A**) Schematic representations of the possible mechanism about the Mn-dhbq crystal structure changes corresponding with the stimulus. (**B**) The length between the two methyl groups for PX, MX, and OX molecules. (**C**) The comparison of PXRD patterns of as-synthesized, activated, and xylene-adsorbed Mn-dhbq samples. (**D**) The thermogravimetric (TG) curves of xylene@Mn-dhbq samples. (**E** to **G**) The adsorption binding sites, π-π stacking (site Is, green dotted lines), hydrogen bonds (site IIs, blue dotted lines), and $C_{methyl}^{\delta-}\cdots Mn(II)^{\delta+}$ dipole-dipole interactions (site IIIs, red dotted lines) between the xylene molecules and the 1D chain of the Mn-dhbq framework as determined by DFT simulations.

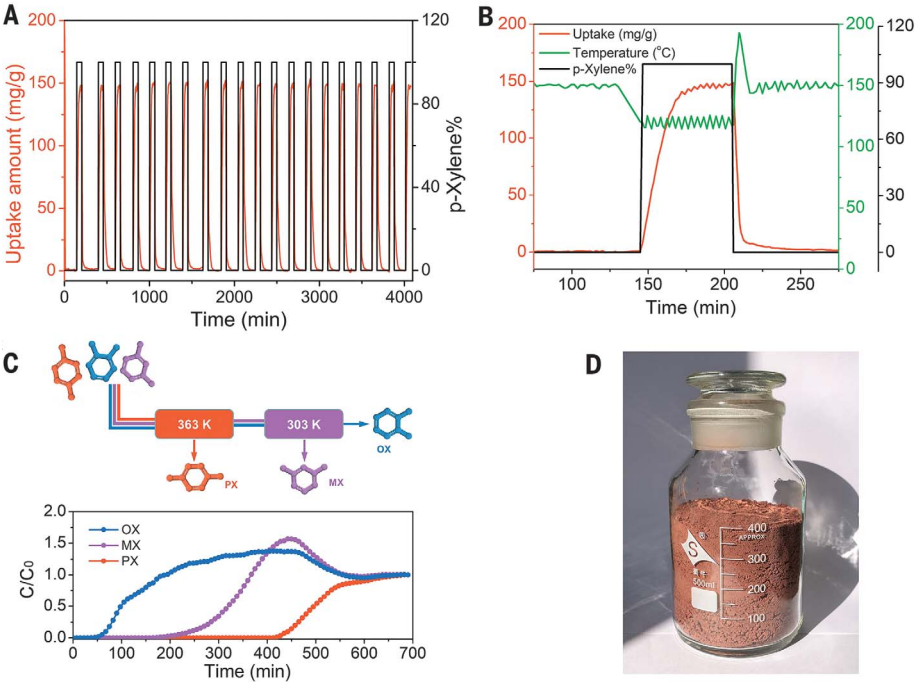


Fig. 4. Recyclability tests, breakthrough separation of xylene ternary mixture, and large-scale synthesis of Mn-dhbq. (**A**) The *p*-xylene adsorption-desorption recyclability test on Mn-dhbq powder samples for 20 consecutive adsorption cycles at 393 K. The orange and black curves indicate the uptake amount of *p*-xylene and the concentration of *p*-xylene combined with N_2 flow, respectively. (**B**) Enlarged image of the first cycle from (A) showing the *p*-xylene adsorption on Mn-dhbq at 393 K and desorption at 423 K under nitrogen flow. The green curve is the temperature of the adsorption chamber. (**C**) Breakthrough curves of an equimolar ternary mixture of all three xylenes (bottom) using Mn-dhbq pellets packed in two columns at 363 and 303 K (top). The partial pressure of xylenes in the feed gas used for the ternary breakthrough experiment is 3.9 kPa. (**D**) Scaled-up synthesis of as-synthesized Mn-dhbq obtained from rapid string and mixing of an aqueous solution of manganese acetate tetrahydrate and an aqueous solution of H_2dhbq ligand.

REFERENCES AND NOTES

1. M. Minceva, A. E. Rodrigues, *AIChE J.* **53**, 138–149 (2007).
2. G. Zhang, Y. Ding, A. Hashem, A. Fakim, N. M. Khashab, *Cell. Rep. Phys. Sci.* **2**, 100470 (2021).
3. D. S. Sholl, R. P. Lively, *Nature* **532**, 435–437 (2016).
4. Y. Yang, P. Bai, X. Guo, *Ind. Eng. Chem. Res.* **56**, 14725–14753 (2017).
5. L. S. Cheng, J. A. Johnson, "Adsorbents with improved mass transfer properties and their use in the adsorptive separation of para-xylene," US Patent 8,609,925 (2013).
6. B. Van de Voorde, B. Bueken, J. Denayer, D. De Vos, *Chem. Soc. Rev.* **43**, 5766–5788 (2014).
7. G.-Q. Guo, H. Chen, Y.-C. Long, *Microporous Mesoporous Mater.* **39**, 149–161 (2000).
8. K. Jie *et al*., *J. Am. Chem. Soc.* **140**, 6921–6930 (2018).
9. N. Sun *et al*., *Chem. Sci.* **10**, 8850–8854 (2019).
10. G. W. Zhang *et al*., *Chem* **6**, 1082–1096 (2020).
11. M. Lusi, L. J. Barbour, *Angew. Chem. Int. Ed.* **51**, 3928–3931 (2012).
12. M. du Plessis, V. I. Nikolayenko, L. J. Barbour, *J. Am. Chem. Soc.* **142**, 4529–4533 (2020).
13. B. Moosa *et al*., *Angew. Chem. Int. Ed.* **59**, 21367–21371 (2020).
14. M. I. Gonzalez *et al*., *J. Am. Chem. Soc.* **140**, 3412–3422 (2018).
15. V. Finsy *et al*., *J. Am. Chem. Soc.* **130**, 7110–7118 (2008).
16. A. Torres-Knoop, R. Krishna, D. Dubbeldam, *Angew. Chem. Int. Ed.* **53**, 7774–7778 (2014).
17. X. Li *et al*., *Nat. Commun.* **11**, 4280 (2020).
18. J. M. Holcroft *et al*., *J. Am. Chem. Soc.* **137**, 5706–5719 (2015).
19. L. Alaerts *et al*., *Angew. Chem. Int. Ed.* **46**, 4293–4297 (2007).
20. L. Yang *et al*., *ACS Appl. Mater. Interfaces* **13**, 41600–41608 (2021).
21. D. D. Zhou *et al*., *Nat. Mater.* **18**, 994–998 (2019).
22. H. Wang *et al*., *Energy Environ. Sci.* **11**, 1226–1231 (2018).
23. P. J. Bereciartua *et al*., *Science* **358**, 1068–1071 (2017).
24. C. Gu *et al*., *Science* **363**, 387–391 (2019).
25. L. Alaerts *et al*., *J. Am. Chem. Soc.* **130**, 14170–14178 (2008).
26. S. Q. Wang *et al*., *Angew. Chem. Int. Ed.* **58**, 6630–6634 (2019).
27. S. Mukherjee *et al*., *Sci. Rep.* **4**, 5761 (2014).
28. X. Cui *et al*., *Nat. Commun.* **11**, 5456 (2020).
29. S. Morikawa, T. Yamada, H. Kitagawa, *Chem. Lett.* **38**, 654–655 (2009).
30. T. Yamada, S. Morikawa, H. Kitagawa, *Bull. Chem. Soc. Jpn.* **83**, 42–48 (2010).
31. R. W. Neuzil, "Aromatic hydrocarbon separation by adsorption," US Patent 3,558,730 (1971).
32. D. M. Polyukhov, A. S. Poryvaev, A. S. Sukhikh, S. A. Gromilov, M. V. Fedin, *ACS Appl. Mater. Interfaces* **13**, 40830–40836 (2021).
33. J. Crank, *The Mathematics of Diffusion* (Oxford Univ. Press, 1975).
34. M. Minceva, A. E. Rodrigues, *Chem. Eng. Res. Des.* **82**, 667–681 (2004).

ACKNOWLEDGMENTS

We thank Y. Liu and J. D. Hu for their help in the graphic work in Fig. 3. **Funding:** This work was supported by the National Natural Science Foundation of China (21722609, 21878260, and 22141001) and the Zhejiang Provincial Natural Science Foundation of China (LR170B060001). The Rutgers team thanks the US Department of Energy for the partial support through grant DE-SC0019902. **Author contributions:** Z.B. and J.L. conceived and designed the project. L.L. synthesized the materials, performed most of the vapor adsorption and breakthrough experiments, and drafted the manuscript. L.G. carried out the molecular simulation and analyzed the binding sites of the guests. D.H.O. collected and analyzed the kinetic adsorption data. S.X. conducted sorption cycling tests. K.W. prepared the shaped pellets for the vapor adsorption isotherms. Z.Z., Q.Y., and Y.Y. discussed the experimental results and gave valuable suggestions. J.L. and Q.R. supervised the experiment and revised the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj7659
Materials and Methods
Figs. S1 to S67
Tables S1 to S6
References (*35*–*60*)
Data S1

Submitted 31 May 2021; resubmitted 12 April 2022
Accepted 15 June 2022
10.1126/science.abj7659

By **Mary O'Reilly**

An artist at heart

Three years into my postdoc, I started to wonder whether I needed a new career plan. After applying for more than two dozen faculty jobs, I hadn't landed a single interview. I felt dejected but not particularly surprised. I was applying in the middle of the 2008 financial crisis, when many universities had instituted hiring freezes and faculty openings were scarce, and my publication record didn't stand out. I could have spent another year or two as a postdoc waiting out the financial storm and building up my CV. But my future husband lived across the country and we were eager to close the gap one way or another. I needed a plan B—and a long dormant dream of becoming an artist began to stir.

I had once considered going to art school but had put that notion to the side when I decided to pursue chemistry as an undergraduate. In the years that followed, I kept up my interest in art by taking drawing and painting classes at night. My family was bursting with mathematicians, computer programmers, and engineers who pursued music and devoured literature in their spare time, so it had felt natural to have my daily life revolve around science, with art as my dreamy lunar companion.

But in the spring after my failed job search, that started to change after an office mate excitedly showed me proofs for a review article. She was wowed by what the journal's scientific illustrator had done with her rudimentary sketches. "That would be such a fun job," I mused.

"I love that I get to combine my scientific and artistic sides."

I decided to test out a new career direction by volunteering to create similar illustrations for my institute's newsletters. I spent my nights and weekends reading scientific papers and thinking about how to illustrate the results. It was a fun task—something that engaged my artistic, creative side and made use of my scientific training. I felt I was perhaps on the right path. But could I make a full-time career work?

Searching online, I tracked down people who had that kind of job. I found many had training through scientific illustration master's degree programs. After living on grad student and postdoc salaries for years, I didn't have enough money saved up for tuition, so I decided to get a certificate in digital design and forge my own path.

It was exciting to find a career that drew on my diverse skill set and would allow me to work as a freelancer from wherever my partner got a job. I did feel a sense of loss as I began to let go of my dream of becoming a faculty member, and I worried I'd be letting down the people and institutions who'd invested time and resources in me. But it helped to remind myself that my new career path wasn't removed from science. I was harnessing my passion for art in the service of science. And some key early projects, which required me to thoroughly understand the research, convinced me that my background was a crucial part of this niche I was carving out for myself.

As I launched my fledgling career, I took an adjunct position teaching chemistry at a university near where we were living. The income gave me breathing room to get my freelance business off the ground, which wasn't easy. I misguidedly spent much of the money I earned from illustrating on marketing to try to drum up business. But I quickly learned that word of mouth and online searches, not advertising, were the best sources of new projects for me. Fortunately, as my client list grew, so too did the referrals. It took more than 3 years, but eventually my business grew sufficiently large that I decided to stop teaching.

I now work as a visual designer at a biomedical research institute, where I spend my days working with researchers to communicate their work visually. I love that I get to combine my scientific and artistic sides and contribute to the dissemination of knowledge to the scientific community. And though I can't in good conscience recommend my long and winding path to this career, I wouldn't change a thing about the stops I made along the way. ■

Mary O'Reilly is a visual designer for Pattern, a design and data visualization group at the Broad Institute.

ILLUSTRATION: ROBERT NEUBECKER

Temperature Gradient System for Mass Spectrometry

ionRocket visualizes hidden compounds in your samples and generates a temperature gradient from ambient up to 600ºC over several minutes. This allows compounds in your samples to be sublimated, vaporized, or pyrolized according to their volatility, and then introduced into the DART gas stream. Data obtained from ionRocket yields another axis of data (time/temperature) beyond that obtained from normal DART-MS (direct analysis in real time-mass spectrometry) analysis. Species desorb in order of their volatility along the temperature gradient and thus are separated in time. The data resembles that of an LC-MS or GC-MS chromatogram, consisting of temperature/time, *m*/*z*, and intensity. This can separate rare from abundant species, making them easier to detect.

BioChromato
For info: +81-4-6623-8382
biochromato.com/analyticalinstruments/ionrocket

Antigen Retrieval System

EZ-Retriever System is a microwave-based antigen retrieval pretreatment system invented by BioGenex. It is used for dewaxing, rehydration, and antigen retrieval of formalin-fixed, paraffin-embedded (FFPE) tissue sections to enhance exposure of antigenic epitopes. This system not only resolves inconsistencies due to different types of antigen retrieval methods by applying different maximum microwave power levels, solution volumes, and heating time, but also facilitates the standardization of dewaxing and antigen protocols for high-quality, reproducible staining.

BioGenex
For info: +1-800-421-4149
biogenex.com/ar-ez-rt-system/?v=7516fd43adaa

Isolation of High-Purity Exosomes

AMS Biotechnology (AMSBIO) is a leading supplier of high-purity exosomes using isolation methods, including traditional differential and gradient ultracentrifugation, immunoaffinity, and size-exclusion chromatography as well as precipitation/low-speed centrifugation. Exosomes are small endosome-derived lipid nanoparticles actively secreted by exocytosis in most living cells. Isolated from diverse cell lines, exosomes play significant roles in diverse pathological conditions such as cancer, and in infectious and neurodegenerative diseases. For labs requiring isolation and sequencing of exosomal microRNA (miRNA) isolation, AMSBIO offers a service to extract RNA from exosomes and subject it to RNA-sequencing using an Illumina platform. In addition to performing basic analysis on the raw data, we can also screen for specific miRNA. AMSBIO also offers analysis of exosomal surface proteins using the MACSPlex Exosome Kit, as well as a top-quality service that provides reliable, reproducible isolation of exosomes from almost any biofluid ranging from plasma and tissue culture media to cerebrospinal fluid. Using this service, your target exosomes are isolated by ultracentrifugation before quantification and quality assessment using a NanoDrop spectrophotometer.

AMS Biotechnology
For info: +1-617-945-5033
www.amsbio.com/exosome-research-services

Automated Barcode Applicator

Porvair Sciences announces eGecko²—an automated barcode application system that provides the ideal high-throughput solution for precisely applying barcode labels onto racks of plates and Petri dishes. Based upon a robust and proven label printer and print engine, eGecko² is a completely integrated printer and application unit, requiring only an electrical supply for operation. A choice of easily exchangeable operating heads allows you to apply labels on both flat and curved surfaces. Programmable Z-axis motion control application macros allow the eGecko² to automatically adjust between plate types (shallow to deep well) and Petri dishes and then position and apply the labels accordingly. Powerful yet highly intuitive software enables label positioning while simultaneously running the eGecko² automation. Using its fully integrated PC, eGecko² enables you to connect to your network laboratory information management system to allow live data transfer and validation of application labels.

Porvair Sciences
For info: +1-856-696-3605
www.kbiosystems.com/label-applicators/l11.htm

Automated Tube Selector

A new adapter kit enables the Ziath Mohawk tube selector to quickly and easily pick and reformat tubes from Azenta Life Sciences Acoustix racks used by the latest generation of Labcyte Echo acoustic liquid dispensers. The Acoustix adapter for Mohawk reduces the effective height of the Mohawk tube selector lid by introducing a clear acrylic plate at a lower height, which easily prevents Acoustix tubes from exiting the inverted rack. This both simplifies and improves the speed of Acoustix tube picking on the Mohawk. In addition, Ziath has developed a release plate that can quickly and easily unlock all 96 tubes in a rack, ready for picking. Using a new Acoustix software plug-in combined with the new adapter and release plate, users can expect fast and easy single-tube cherry picking and reformatting from Acoustix racks, with automatic confirmation of pick-list integrity.

Ziath
For info: +1-858-880-6920
ziath.com/products/tube-devices/mohawk

Liquid Dispenser Platforms

DISPENDIX announces the release of I.DOT HT Non-Contact Dispenser and the L.DROP Liquid Handler. The I.DOT HT builds on the core technology of the I.DOT, expanding the source-well layout to a 384-well format to support management of compound libraries at large pharmaceutical and small biotechnology companies. The L.DROP couples sensor technology and web-based architecture to provide a more dynamic instrument that is capable of scaling protocols across instruments and even laboratories without tedious optimization. Features such as automated deck calibration, web-based protocol fitting, and closed-loop pipetting promise to revolutionize the way researchers and automation engineers leverage liquid-handling technologies.

DISPENDIX
For info: +1-833-235-5465
www.dispendix.com

Electronically submit your new product description or product literature information! Go to www.science.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.