Science

$15
14 JANUARY 2022
science.org

◣◢AAAS

DISPERSAL
DECLINE

Plants at risk from losses
of seed dispersers p. 210

TRILLIONS OF MICROBES
ONE ESSAY

The **NOSTER** *Science* **Microbiome Prize is an international prize that rewards innovative research by investigators, under the age of 35, who are working on the functional attributes of the microbiota. The research can include any organism that has potential to contribute to our understanding of human or veterinary health and disease, or to guide therapeutic interventions. The winner and finalists will be chosen by a committee of independent scientist, chaired by a senior editor of *Science*. The top prize includes a complimentary membership to AAAS, an online subscription to *Science*, and $25,000 (USD). Submit your research essay today.**

Eran Blacher, Ph.D.
2021 Winner

NOSTER | Science
MICROBIOME
PRIZE

Apply by 1/24/22 at www.science.org/noster
Sponsored by Noster, Inc

CONTENTS

14 JANUARY 2022 • VOLUME 375 • ISSUE 6577

134

NEWS

CONTENTS

RESEARCH

221
Membrane pores are formed by bacterial gasdermin

ON THE COVER

A blue-throated barbet (*Psilopogon asiaticus*) eats a fig in southern Bhutan, dispersing the seeds within. Seed-dispersing animals help fleshy-fruited plants adapt to climate change by moving seeds to suitable habitats. However, many seed dispersers are in decline, endangered, or extinct. Bird and mammal population losses have caused an estimated 60% decline in seed dispersal, limiting plants' ability to track climate change. See page 210.
Photo: Christian Ziegler

When hyping technology is a crime

The rise and fall of Elizabeth Holmes, former chief executive officer (CEO) of the high-flying, privately held, American medical diagnostics company Theranos, has riveted the public for years. A bestseller, a documentary, and a seemingly endless stream of news stories have chronicled the drama, which came to a climax last week when a jury decided that Holmes committed fraud. Although a verdict has been rendered, it's worth examining how the culture of science innovation contributed to the problem. For too long, fledgling companies promoting technological and scientific advances have relied too much on style and not enough on substance.

An exciting and robust technology is generally only part of the picture in creating a successful life science business. Investors—especially venture capitalists in Silicon Valley—are also looking for a compelling CEO who can dazzle other investors and eventually Wall Street. It's often said that to succeed, a venture-backed company needs a CEO who is a "force of nature"—someone who can raise money for a promised new technology even when the supporting data are not great. This person should be someone who the investors believe will "get there or die trying" and who has the ability to "fake it till you make it" to keep the company going through the inevitable setbacks along the way. None of these traits have anything to do with developing a robust technology that saves lives. Pushed too far, the exaggerations become fraud, and the Holmes case shows that can be a crime.

As venture-backed companies grow, investors usually insist on "milestones" that show the steady improvement of the technology. The temptation to misrepresent the progress increases with each step. If the technology is not actually advancing, then the executives in the company are committing greater and greater acts of fraud. There is also a burden on the investors to carry out due diligence on the claims, but they only have financial and not criminal liability. When attention is too focused on charisma and a certain amount of fakery is considered acceptable, reality can be overcome by hype.

That is what happened with Theranos. Holmes was more obsessed with her personal impression and less so about whether the company delivered on its promise to provide small, automated devices that required very small amounts of blood to perform numerous diagnostic tests. Her boyfriend and business associate Ramesh "Sunny" Balwani—who awaits his own trial—presumably provided instructions on her personal style and conduct, but nothing about how to make the technology work. Maybe reading some scientific papers about diagnostic technologies would have been a good thing. Instead, investors happily followed the 19-year-old Stanford dropout (with no experience) down an entrepreneurial path paved with dishonesty, not reality.

With information technology companies, the stakes are not so high. If tech CEOs overhype their shiny new idea, venture capitalists will merely lose money that they knew full well was at substantial risk. But fakery involving new medical tests can lead to life-threatening harm. Interestingly, the jury found Holmes guilty of defrauding investors, but not defrauding patients. The latter verdict may have been influenced by the fact that physicians who used the Theranos blood tests followed up with other lab tests, thereby limiting the potential damage to people's lives. The implications of this for the commercialization of science are enormous. Being optimistic that the science might work is one thing, saying it works when it doesn't is another. In Theranos's case the distinction was clear-cut, but the efficacy of many other technologies—particularly in the life sciences—is not as easily deduced. If it works in cells, will it work in animals? If it works in animals, will it work in humans? Can you produce it at scale?

"Get there or die trying" sounds snazzy, but the Theranos case is a reminder that more than money can be at stake. What happened to Holmes should be a warning for future start-ups. Instead of forces of nature, how about companies led by highly accomplished scientists who give dull and boring PowerPoint presentations full of outstanding data?

"When attention is too focused on charisma... reality can be overcome by hype."

H. Holden Thorp
Editor-in-Chief, *Science* journals.
hthorp@aaas.org;
@hholdenthorp

–H. Holden Thorp*

PHOTO: CAMERON DAVIDSON

Science ▲AAAS

Lessons learned from leading NIH

Having just stepped down from serving three presidents over 12 years as director of the US National Institutes of Health (NIH), I am deeply grateful for the opportunity to have led this noble enterprise and to work with amazing people. Astounding progress has been made on many fronts, including advances that I never dreamed that I would see in my lifetime—let alone during my time as NIH director.

Who would have thought that understanding biology at the single-cell level would move forward so rapidly? Or that immunotherapy would emerge as an extremely promising strategy for treating cancers? Few would have predicted that gene editing would be possible in nearly every lab, paving the way for cures of disorders like sickle cell disease. Even more difficult to foresee was the worst global pandemic in more than a century, COVID-19, followed by the unprecedented achievement of using a new messenger RNA platform to develop vaccines to combat it in less than a year.

Thanks to an engaged US Congress, NIH has experienced a remarkable period of budget growth since 2016, which has made it more feasible to initiate new projects. Yet questions remain about whether this supportive trajectory can be sustained. Equally concerning is a hyperpartisan and divisive atmosphere that has taken a toll on the credibility of science. Objective truth is under attack. For COVID-19, this has taken the form of unsubstantiated claims about the origins of the disease-causing coronavirus, rejection of evidence-based information on prevention with masks and vaccines, and embrace of highly questionable, potentially harmful "treatments."

What have I learned during my time at this agency's helm? With the aim of making life easier for NIH's next director, there are several lessons that rise to the top. Hypothesis-driven research is the bedrock of NIH's success. Thus, one should prioritize support of risk-taking research and pay less attention to preliminary data and more to the potential importance of a premise. Also, it is key to survey the biomedical landscape for areas poised for bold advances and then zero in on those exceptional opportunities. During my directorship, teams were organized to pursue such opportunities, including the National Center for Advancing Translational Sciences, the

> **"...NIH exists ...to offer hope...and solutions to illnesses that cause so much suffering..."**

"All of Us" Research Program, the Human Health and Heredity in Africa program, The Cancer Moonshot, and the Brain Research Through Advancing Innovative Neurotechnologies Initiative. I found that nimbleness matters. NIH has not always been known for rapid responsiveness, but in recent years has learned to move more nimbly in times of urgent need. Public-private partnerships enable research to proceed much faster, as evidenced by NIH's Accelerating Medicines Partnership, and the Accelerating COVID-19 Therapeutic Interventions and Vaccines and Rapid Acceleration of Diagnostics initiatives. Establishment of an Advanced Research Projects Agency for Health within NIH would serve powerfully to maintain this momentum. It is also crucial to remember that NIH serves the people. Consequently, it is essential to cultivate relationships with a broad cross-section of society, including Congress, patient groups, and colleagues in all sectors. And it is vital that the agency play a strong role in encouraging the next generation of researchers. Aspiring researchers are biomedical science's most important resource, and every effort should be made to provide them with welcoming "on ramps" to productive careers. At the heart of such encouragement is recognizing the value of diversity. NIH has made progress in this area through efforts like UNITE and the Faculty Institutional Recruitment for Sustainable Transformation, but long-term commitment is needed to achieve sustained diversity and equity.

Biomedical research faces many challenges, but I am confident that the next NIH director will meet them. COVID-19 has unquestionably been a harsh challenge, and the global importance of investing in biomedical research has never been clearer. I leave NIH in the hands of a highly capable acting director and a remarkable group of institute and center directors, most of whom I am proud to say I recruited.

I am fundamentally a person filled with hope. Indeed, the very reason that NIH exists is to offer hope to those seeking answers to the mysteries of life and solutions to illnesses that cause so much suffering and premature death. I am honored to have played a role in providing such hope for the last 12 years and look forward to supporting those who will succeed me in pursuing NIH's never-ending quest to turn biomedical discoveries into better health for all humankind.

–Francis S. Collins
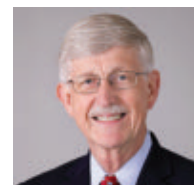
Francis S. Collins was director of the US National Institutes of Health, Bethesda, MD, USA, from 2009 to 2021. francis.collins @nih.gov

IN BRIEF

Edited by **Lila Guterman**

ECOLOGY

Huge icefish colony found

Icefish guard nests along the Weddell Sea's floor.

S cientists aboard an Antarctic research cruise have discovered the most extensive breeding colony of fish anywhere. In February 2021, while towing video cameras and other instruments close to the sea floor in the Weddell Sea, the *RV Polarstern* came upon thousands of 75-centimeter-wide nests, each occupied by a single adult icefish (*Neopagetopsis ionah*) and up to 2100 eggs. "It was really an amazing sight," says deep-sea biologist Autun Purser of the Alfred Wegener Institute, who led the research cruise's underwater imaging, reported in *Current Biology* this week. Including subsequent tows, the team saw 16,160 closely packed fish nests, 76% of which were guarded by solitary males. Adult icefish may use currents to find the spawning grounds, where the water above is rich in zooplankton that their offspring can eat, Purser says. Assuming a similar density of nests in the areas between the survey lines, the researchers estimate that about 60 million nests cover roughly 240 square kilometers. The icefish and their eggs dominate the area's biomass, which the authors describe as "exceptionally high" for the Antarctic sea floor. The researchers say the colony provides a new reason to create a marine protected area in the Weddell Sea, a unique and largely undisturbed ecosystem.

Foreign funding disclosure

RESEARCH SECURITY | U.S. research agencies have 120 days to develop uniform policies describing the outside sources of funding that scientists must disclose when they apply for federal grants. The deadline was set in a 4 January directive from the White House Office of Science and Technology Policy (OSTP) aimed at protecting federally funded research from attempted theft by some foreign governments. In recent years, the Department of Justice has prosecuted some two dozen academics for failing to disclose financial ties to China, a move critics say has chilled research collaborations and criminalized minor violations of often confusing federal rules. Research groups welcomed OSTP's 34-page report, but say they wish it had specified what kinds of foreign collaborations might get a scientist in trouble. "OSTP intends to address [such questions] in the future," says its director, Eric Lander.

And one federal agency, the Defense Advanced Research Projects Agency, already has begun to give top-ranked proposals a second look, based on the foreign ties of the scientist submitting the proposal, before making any award.

Chinese tire of 'zero COVID' policy

COVID-19 | As the 13 million residents of the city of Xi'an endure a third week of lockdown after a December 2021 outbreak, many Chinese are losing patience with the country's "zero COVID" strategy. People expressed outrage on social media after a graphic 4 January post about a woman in her last month of pregnancy whose admission to a Xi'an hospital for abdominal pain was delayed because her most recent COVID-19 test was no longer valid; she had a stillbirth at the hospital entrance, blood pooling at her feet. (The original post has disappeared.) Authorities say the tough measures have nearly ended Xi'an's outbreak, the worst in China since the disease broke out in Wuhan, in early 2020. Since 9 December 2021, Xi'an has logged more than 2000 Delta variant infections, but on 10 January, it reported just 13 new cases. China has now detected its first community cases of the Omicron variant, sending more cities into lockdown and raising concerns about the impact on the Winter Olympic Games, starting on 4 February in Beijing.

Ticks with Lyme overwinter better

ECOLOGY | More ticks carrying the bacterium that causes Lyme disease survived the winter in Nova Scotia than did uninfected ones in a new study of 600 of the arachnids kept outside in small vials. "Winter conditions may favor the ability of infected ticks to find hosts and continue to spread disease," says Laura Ferguson, an eco-immunologist at Dalhousie University. And infected ticks were more active during fluctuating wintry temperatures in the lab, Ferguson and colleagues reported at the annual meeting of the Society of Integrative and Comparative Biology in Phoenix last week. The finding suggests the variable winter conditions brought on by climate change could boost the odds people will encounter Lyme-infected ticks.

New research cannabis imminent

BIOMEDICINE | Two companies are preparing to ship cannabis to academic and commercial clients for research use, having won long-awaited registrations that break the University of Mississippi's 54-year monopoly on U.S. research cannabis

Surgeons examine the genetically engineered pig heart transplanted into a person last week.

BIOMEDICINE

Man receives pig heart

Surgeons last week performed the first pig-to-human heart transplant on a 57-year-old man with a life-threatening heart condition who was ineligible for a human donor organ. The heart, provided by the biotechnology company Revivicor, came from a pig genetically engineered to prevent its organs from prompting immune rejection by the human body. The patient was doing well 4 days after the 7 January transplant, according to surgeons at the University of Maryland Medical Center, but rejection remained a risk. If the strategy proves successful and safe long-term, pig organs could offer a long-hoped-for supply for people awaiting transplants.

production. Groff North America, based in Red Lion, Pennsylvania, harvested its first crop on 30 December 2021 and expects to ship cannabis preparations in the first quarter of this year to clients including universities and biopharmaceutical companies. The Biopharmaceutical Research Company (BRC), based in Castroville, California, plans to harvest its first crop intended for research customers in about 8 weeks. The company is also developing cannabis research programs with Washington State University and the University of California, Davis. Groff North America and BRC won bulk manufacturing registrations from the U.S. Drug Enforcement Agency in May and June 2021, respectively.

Dutch science head takes office

PEOPLE | Theoretical physicist Robbert Dijkgraaf, who since 2012 has headed the Institute for Advanced Study (IAS) in Princeton, New Jersey, was sworn in as minister of education, culture, and

science in the Netherlands on 10 January. Dijkgraaf, a popular science communicator, newspaper columnist, and talk show guest in his home country, had called for an increase in government funding for science, currently some €6.2 billion annually. He will get his wish: The new center-right coalition of four parties has agreed to launch a €5 billion fund for R&D, to be spent over the next decade. IAS—which was home to Albert Einstein and many other great scientists—announced in November 2021 that Dijkgraaf will be succeeded by medieval historian David Nirenberg, dean of the University of Chicago Divinity School, in July.

China pioneers small reactor

ENERGY | A Chinese utility last month connected the world's first commercial-scale small nuclear power plant of its kind to the electric grid. Several nations are pursuing small modular reactors—compact units that can be mass produced—because they

Engineers monitor James Webb Space Telescope operations.

ASTROPHYSICS

Webb telescope fully formed

The James Webb Space Telescope, the largest space observatory ever launched, completed most of its delicate in-space deployments last week after unfolding two wings of its 6.5-meter main mirror. Webb has now executed 85% of several hundred mechanical actions that could jeopardize the mission were any to fail. Next, operators will fine-tune the positions of the main mirror's 18 gold-plated segments with 126 tiny motors to form a single reflecting surface; then, the team will calibrate Webb's four instruments, which must be cooled below −230°C to see in the infrared. Expected this summer is a first image, one that will demonstrate Webb's unprecedented capability to peer into the early universe and other worlds—then the science begins.

are cheaper to build and theoretically safer. At the Shidao Bay Nuclear Power Plant in eastern China, designers opted for a "pebble bed" design, in which nuclear fuel is encased in spherical graphite "pebbles" the size of tennis balls that are packed into a cylindrical vessel like gumballs in a jar. Circulating helium is heated by the pebbles to 750°C, and the hot gas produces steam. An identical pair of these reactors will jointly generate 210 megawatts of electricity. Developers are now planning to link six reactor modules together in a 650-megawatt plant. U.S. regulators have yet to approve a leading small-reactor design from the company NuScale Power because of concerns over the reliability of its passive safety features.

Ocean warming sets record

CLIMATE | The upper reaches of the world's oceans were hotter than ever in 2021, setting a record for the sixth straight year. The shallowest 2000 meters of the oceans last year absorbed 14 more zettajoules (10^{21} joules) of heat than in 2020, an injection of energy equivalent to nearly 30 times as much energy as humans use in a single year, scientists reported this week in *Advances in Atmospheric Sciences*. The ocean acts as a buffer against rising temperatures, absorbing an estimated 90% of the extra heat trapped by global warming. Surface air temperatures, more mercurial than the ocean, didn't set a record

in 2021—it was merely the fifth warmest year, and between 1.1°C and 1.2°C above the late 19th century's average, the European Union's Copernicus Climate Change Service announced this week. Despite economic slowdowns triggered by the pandemic, two key greenhouse gases rose to record levels in the atmosphere, with carbon dioxide at 414 parts per million and methane at 1876 parts per billion, according to the service.

NASA names chief scientist

CLIMATE POLICY | Katherine Calvin, a researcher who studies the impacts of global warming, was appointed this week to be NASA's new chief scientist and senior climate adviser. Calvin has worked since 2008 at the Pacific Northwest National Laboratory, modeling how climate change will alter farming, water use, and other aspects of human society. Calvin will advise Bill Nelson, NASA's administrator, as the agency grapples with adapting to climate change, including planning the long-term future of the Kennedy Space Center in Florida and other prominent space facilities on coastlines vulnerable to sea level rise. (The chief scientist position does not directly oversee NASA's scientific missions, including those focused on climate; those are managed by Associate Administrator Thomas Zurbuchen.)

Pill curbs but doesn't erase yaws

INFECTIOUS DISEASES | Eradicating yaws—a painful and disfiguring bacterial disease—may be harder than scientists hoped, a study published last week in *The New England Journal of Medicine* suggests. The World Health Organization aims to eradicate yaws by 2030; its strategy calls for mass treatment of communities using a single dose of the inexpensive oral antibiotic azithromycin, followed by targeted treatment for people who develop the disease's hallmark skin ulcers and their contacts. In the trial, involving 56,000 people in Papua New Guinea, that regimen reduced yaws prevalence by 65%. An alternative, intensified treatment—three communitywide doses of azithromycin given 6 months apart—did markedly better, lowering prevalence by 91%. But it didn't wipe out the disease, as the scientists had hoped, and a few cases of antibiotic resistance resulted, raising concern.

S SCIENCE.ORG/NEWS
Read more news from *Science* online.

Images of fields filled with funeral pyres, like these in New Delhi in April 2021, highlighted the massive number of COVID-19 deaths in India.

COVID-19

India's pandemic toll far exceeds official count

New analysis bolsters idea that country's seemingly low death rate was misleading

By **Jon Cohen**

India, from the earliest days of the pandemic, reported far fewer COVID-19 deaths than expected given the toll elsewhere—an apparent death "paradox" that some believed was real and others thought would prove illusory. Now, a prominent epidemiologist who contended the country really had been spared the worst of COVID-19 has led a rigorous new analysis of available mortality data and concluded he "got it wrong."

India has "substantially greater" COVID-19 deaths than official reports suggest, says Prabhat Jha of the University of Toronto—close to 3 million, which is more than six times higher than the government has acknowledged and the largest number of any country. If true, the finding could prompt scrutiny of other countries with anomalously low death rates and would dramatically push up the current worldwide pandemic total, estimated by the World Health Organization (WHO) at some 5.45 million people.

"I think it does call for a recalibration of the global numbers plus saying, 'What the heck is going on in India?'" says Jha, whose team published the new India analysis online

last week in *Science*. And India's suffering could be far from over—the Omicron variant of the coronavirus is surging there.

At the end of 2021, India reported about 480,000 deaths from SARS-CoV-2 infections. That's 340 COVID-19 deaths per million—about one-seventh the per capita COVID-19 mortality tallied in the United States. Jha's own early analysis supported the apparently low mortality rate from COVID-19, but he and his colleagues have now probed more deeply. They tapped data from an independent polling agency that surveyed nearly 140,000 people across the vast country by telephone, asking whether anyone in each household had died from COVID-19. They also analyzed government reports from hospitals and similar facilities and looked at officially registered deaths. The result: a much higher estimate—between 2300 and 2500 deaths per million by September 2021, comparable to the rate in the United States and pointing to a much higher total death toll, because India has four times as many people.

Jha says his early, low estimate was based on the first wave of infections in the fall of 2020, which may have been less deadly than the Delta variant that drove India's massive surge in spring of 2021. He also fo-

cused on the large cities, where death rates may have been lower than in the countryside. And death registration in the country had been spotty even before the pandemic. But those factors can't be the whole story, he says. "There must be other things that we still don't understand."

One, he says, is politics: He thinks the administration of Indian Prime Minister Narendra Modi clouded the true picture of the pandemic. "The Indian government very much is trying to suppress the numbers in the way that they coded the COVID deaths," Jha says. He and others also fault the government for not releasing data from what's known as the Sample Registration System (SRS), which routinely surveys 1% of India's population to track births and deaths. "I think the political pressures were such that they said, 'Anything that's going to come out is going to be embarrassing.'"

Ramanan Laxminarayan, a Princeton University epidemiologist and economist, doesn't see the undercount as entirely deliberate. He notes that SRS data haven't been released since 2018, before the pandemic, so that suspension may just reflect a disorganized system. Almost every country undercounts COVID-19 mortality, he adds—"I think all governments want to

downplay the degree of deaths." (Three Indian government sources that *Science* contacted did not reply to requests for comments on the new study.)

The new estimates for India come as little surprise to Laxminarayan. "My starting point is that unless you can tell me why India is different, I'm going to assume that India is the same as any other country," he says. "I don't believe in exceptionalism of any kind unless it's well justified." His team last month published a study in *The Lancet* that focused on the Indian district of Chennai and concluded that reported deaths "greatly underestimated pandemic-associated mortality."

Virologist Shahid Jameel of Ashoka University says the countrywide estimates by Jha's team are also "in broad agreement" with two other independent studies that examined a similar time frame. "India paid a heavy price for not having good real-time data on deaths, especially during the first wave. That led to complacency and a terrible toll in the second wave," Jameel says.

The work nicely triangulates data from different sources, each of which has its own limitations, says Samira Asma, a WHO assistant director general who works on data and analysis. "The study design is robust," Asma says. "Countries can learn from this approach to ... produce country-specific estimates." WHO is now updating its estimates of excess deaths caused by COVID-19 and plans to release them soon, she says.

A worldwide comparison of all-cause mortality before and during the pandemic, published in *eLife* 6 months ago, suggests undercounting is widespread. Russia had 4.5 times more deaths than normal, far beyond its official COVID-19 tally, and the trend has continued, the researchers recently tweeted. Tajikistan, Nicaragua, Uzbekistan, Belarus, and Egypt also had profound disconnects.

Omicron has begun to run riot in India, and Jha warns the country should not count on it causing milder disease than earlier variants, as data from other countries suggest it might. "I would be really careful about those assumptions because they're based on selected populations that you can't take from South Africa or the U.K. or Canada to India," he says.

And Jha is wary of "wishful thinking" that the prior high levels of infections in India and wide-scale vaccination will create population level immunity and reduce severe disease from the variant. "We just don't know enough about how these different variants behave in immunized populations," he says.

Spoken like a researcher who doesn't want to get it wrong again. ∎

COMPUTER SCIENCE

Artificial intelligence unmasks anonymous chess players

Software that identifies unique styles poses privacy risks

By **Matthew Hutson**

Think your bishop's opening, queen's gambit, and pawn play are unique? A new artificial intelligence (AI) algorithm has got your chess style pegged.

AI software can already identify people by their voices or handwriting. Now, an AI has shown it can tag people based on their chess-playing behavior, an advance in the field of "stylometrics" that could help computers be better chess teachers or more humanlike in their game play. Alarmingly, the system could also be used to help identify and track people who think their online behavior is anonymous.

"Privacy threats are growing rapidly," says Alexandra Wood, a lawyer at the Berkman Klein Center for Internet & Society at Harvard University. She says studies like this one, when conducted responsibly, are useful because they "shed light on a significant mode of privacy loss."

Chess-playing software, such as Deep Blue and AlphaZero, has long been superhuman. But Ashton Anderson, a computer scientist at the University of Toronto and principal investigator of the new project, says the chess engines play almost an "alien style" that isn't very instructive for those seeking to learn or improve their skills. They'd do better to tailor their advice to individual players. But first they'd need to capture a player's unique form.

To design and train their AI, the researchers tapped an ample resource: more than 50 million human games played on the Lichess website. They collected games by players who had played at least 1000 times and sampled sequences of up to 32 moves from those games. They coded each move and fed them into a neural network that represented each game as a point in multidimensional space, so that each player's games formed a cluster of points. The network was trained to maximize the density of each player's cluster and the distance between those of different players.

That required the system to recognize what was distinctive about each player's style.

The researchers tested the system by seeing how well it distinguished one player from another. They gave the system 100 games from each of about 3000 known players, and 100 fresh games from a mystery player. To make the task harder, they hid the first 15 moves of each game. The system looked for the best match and identified the mystery player 86% of the time, the researchers reported last month at the Conference on Neural Information Processing Systems (NeurIPS). "We didn't quite believe the results," says Reid McIlroy-Young, a student in Anderson's lab and the paper's primary author. A non-AI method was only 28% accurate.

"The work is really cool," says Noam Brown, a research scientist at Meta (the parent company of Facebook) who has developed superhuman poker bots. He looks forward to chess bots that mimic Magnus Carlsen, the reigning world champion, and says style-aware AI could transform other computer interactions. "There's a lot of interest in chatbots, where you can have a chatbot that would speak in the style of Albert Einstein or something," he says.

The researchers are aware of the privacy risks posed by the system, which could be used to unmask anonymous chess players online. With tweaks, McIlroy-Young says, it could do the same for poker. And in theory, they say, given the right data sets, such systems could identify people based on the quirks of their driving or the timing and location of their cellphone use.

NeurIPS organizers found the study technically impressive but ethically fraught, and accepted it on the condition that the researchers elaborate on the privacy risks. (It could be "of interest to marketers [and] law enforcement," one reviewer commented.) Anderson says they've decided, for now, not to release the software code. ∎

Matthew Hutson is a journalist in New York City.

In HIV infection, a DNA copy of its genome (dark blue in illustration) integrates with human DNA (teal). In people on antiretrovirals for years, these "proviruses" can become trapped in quiescent stretches.
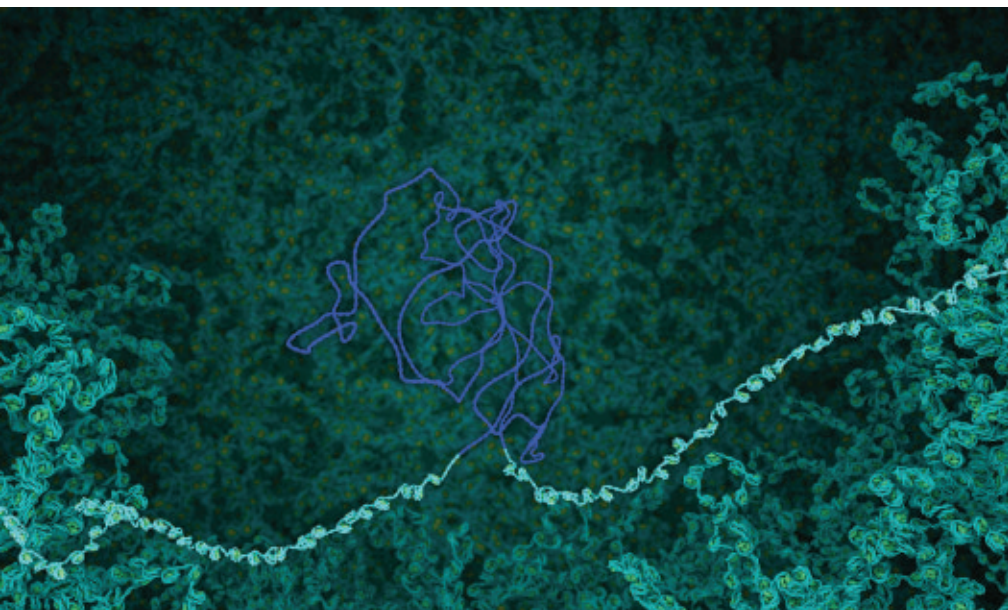
BIOMEDICINE

Mapping where HIV hides suggests cure strategy

Long-term antiretrovirals may corner viral genomes in inactive regions of DNA

By **Jon Cohen**

An HIV infection remains maddeningly difficult to cure because the virus is so good at hiding out. Yes, antiretroviral (ARV) drugs can control an infection, but HIV integrates its genes into human chromosomes, evading both drugs and the immune system. Now, a research group studying a handful of HIV-infected people who have been on treatment for at least 9 years has made a tantalizing discovery: The integrated HIV genomes, or proviruses, that persist the longest are increasingly confined to largely inactive stretches of host DNA, perhaps stymieing production of new viruses—and opening new avenues for cure research.

In earlier work the team had found the same "blocked and locked" integration phenomenon in rare HIV-infected individuals who remain healthy for decades without treatment. Together, the findings raise the possibility of "a peaceful coexistence between HIV and humans," proposes Mathias Lichterfeld, an infectious disease clinician at Brigham and Women's Hospital who led the new analysis.

The group believes long-treated, infected people with this signature integration "landscape" could consider stopping treatment to see whether their immune systems can check any remaining virus production, a strategy they hope to test in a volunteer soon.

The work, published in *Cell* this week, "provides a road map to a cure" for HIV infections, says Steven Deeks, an HIV clinician at the University of California, San Francisco, who was not part of the study but has collaborated with Lichterfeld and his colleagues. Others are more circumspect. "It's good to be cautious until this is done on hundreds of [HIV-infected people]," says Mary Kearney, an HIV researcher at the National Cancer Institute. "But this is a great start."

Over the past 15 years, researchers have tried to drain the reservoirs of HIV hiding in the chromosomes of infected people with drugs meant to prod the proviruses into producing new viruses. The white blood cells that harbor the active proviruses then either self-destruct or become easy prey for other immune warriors. But these "shock-and-kill" strategies have made little headway.

The new strategy builds on a landmark study in which Lichterfeld, Xu Yu of the Ragon Institute of MGH, MIT and Harvard, and other scientists examined "elite controllers," the tiny group of untreated HIV-infected people—less than 0.5%—who live with the virus for decades and suffer no obvious harm from it. In these rare cases, they reported in 2020 in *Nature*, the proviruses tended to cluster in regions of chromosomes that lack genes or harbor largely inactive genes that code for a type of protein known as a zinc finger (ZNF)—which, intriguingly, evolved to repress ancient retroviruses. In both kinds of regions, the DNA is more tightly packed than elsewhere in the human genome, making proviruses less accessible to factors that drive transcription.

In the 15 December 2021 issue of *Science Translational Medicine* (*STM*), Yu and Lichterfeld—who are married—followed up on that finding by proposing how these unusual, virus-hostile landscapes might evolve in elite controllers. (Deeks was a co-author.) Researchers have long thought latent proviruses can form large, effectively invisible cellular reservoirs. But the *STM* study suggests HIV integrated into regions with active genes is never entirely invisible. Instead infected cells produce a trickle of new HIV and, as a result, get eliminated.

"The reservoir has always been described as transcriptionally silent and stable, but we actually find it's transcriptionally active to a large extent," Lichterfeld says. The team concluded that elite controllers have special immune responses and other murky mechanisms that speed the elimination of this active reservoir, resulting in what they described as a "skeleton reservoir" dominated by blocked and locked proviruses.

In the latest work, Lichterfeld's and Yu's team examined 1270 proviruses detected in the blood of six people at different points during prolonged HIV treatment. The group found that in three of the people, intact HIV genes gradually accumulated in human gene deserts and the quiescent ZNF genes. "It's a little bit like a chess game: There are only a few spots left where the king can still go," Lichterfeld says. The result: an integration landscape increasingly similar to the one seen in the elite controllers.

Independently, Lillian Cohn, an immunologist at the Fred Hutchinson Cancer Research Center, made a similar finding:

In people on ARVs for between 4 and 21 years, HIV proviruses were preferentially integrated in certain ZNF genes. Why viral DNA persists in those genes is a mystery that demands further study, some scientists suggest; it may be an accident, or the actions of the ZNF proteins may play a role.

What's key in the new study, Cohn notes, is showing that HIV treatment, and not just the unusual immune abilities of elite controllers, can drive the virus into those quiescent regions. "It doesn't necessarily need to be that people are somehow elite and special, but, rather, we might be able to induce this same phenotype in other people," she says.

How to do that remains an open question, Cohn notes. Researchers have floated several ideas for accelerating the block-and-lock process, including using drugs that target the proviral genes and gum up transcription machinery or therapeutic vaccines that speed elimination of the transcribing proviruses. Others hope that long-term treatment with standard ARVs will be enough.

Yu and Lichterfeld say a participant in one of their studies, a person treated with ARVs for more than 2 decades who has a reservoir landscape resembling an elite controller, has agreed to stop treatment to test their hypothesis. And last month at a meeting in Boston on how to control HIV without ARVs held by amfAR, a foundation that supports HIV/AIDS research, the two dozen attendees decided to seek funding for a new collaboration, headed by Yu. It hopes to enroll large numbers of people who have been on ARVs for decades to examine their HIV integration landscapes and find more candidates for treatment interruption studies.

Dawn Averitt, who recently joined a pilot study to have her proviruses examined by Yu and Lichterfeld, says even if the analysis suggests she is a good candidate to stop treatment, she's scared—the drugs have suppressed her virus for more than 20 years. "It's nerve wracking," says Averitt, who started a nonprofit, the Well Project, to help women living with the virus and later founded the Women's Research Initiative on HIV/AIDS. "The devil you know is better than the devil you don't, right?"

Still, Averitt says if invited, she'd probably join the study, primarily to help others. "I figured out how to do the dance with these drugs, but I really care what it means for all of us," she says. "Imagine the hope, imagine the possibility, of being able to say, 'Worry about keeping it under control now, don't worry about what this means forever.'" ■

PHOTO: DONAL HUSNI/NURPHOTO/GETTY IMAGES

Indonesia's research reform triggers layoffs and protests

A new government agency takes control of the country's science—including the venerable Eijkman Institute

By **Dyna Rochmyaningsih**

On 3 January, Isabella Apriyana grabbed her phone to take a picture of her lab bench and post it on Twitter. "A gloomy Monday morning in the beginning of the year," she tweeted. Apriyana, a research assistant who helped prominent geneticist Herawati Sudoyo map the genomes of Indigenous groups across Indonesia, had just lost her job at the Eijkman Institute for Molecular Biology, along with 112 other people—about 70% of Eijkman's staff. The institute itself, a flagship of Indonesian science with roots in the colonial era, had ceased to exist as an independent lab.

Eijkman, which focuses on genetic disorders, population genetics, and tropical and emerging diseases, has been absorbed into Indonesia's National Research and Innovation Agency (BRIN), established last year to streamline Indonesian science (*Science*, 30 April 2021, p. 449). So far, BRIN has swallowed up 33 research agencies in fields as diverse as archaeology, botany, meteorology, and astronomy, including the entire Indonesian Institute of Sciences (LIPI). Despite protests at Eijkman and elsewhere, hundreds of researchers, technicians, and assistants have lost their jobs because they did not have contracts as civil servants, and BRIN won't offer them such contracts now.

"This is an extraordinary setback for Indonesian science," says Satryo Brodjonegoro, head of the Indonesian Academy of Sciences. He suspects the government is centralizing the research institutes in part to strengthen its control over them. Brodjonegoro says the exodus is disruptive for research and should have been avoided. Science is teamwork, he says: "We can't just break up the research team."

For Eijkman, the merger is the end of an era. Founded in 1888 and housed in a handsome building in central Jakarta, the institute was named after its first director, Dutch pathologist Christiaan Eijkman, who won a Nobel Prize in 1929 for research that led to the discovery of vitamin B1. The institute closed in the 1960s but was resurrected in 1992 by research minister Bacharuddin Jusuf Habibie, an engineer who later became the country's president. Habibie realized "Indonesia needed to have a capacity in basic science," says Sudoyo, one of the first scientists hired after Eijkman reopened.

Laksana Tri Handoko, head of BRIN, says integrating Eijkman proved problematic

The Eijkman Institute, founded in 1888, will lose its independence and its building in central Jakarta, Indonesia.

because it had hired too many people on temporary contracts, which BRIN does not recognize. In addition, because Handoko wants to increase the share of Indonesian scientists who have a Ph.D., Eijkman researchers with a bachelor's or master's degree were offered a chance to stay on if they entered a university Ph.D. program. But sources at the institute say BRIN only gave them 3 months to do so—and many did not succeed. Eijkman's research will move to BRIN headquarters in Cibinong, some 50 kilometers south of Jakarta. Some other activities will be transferred to the Ministry of Health, which will also inherit the building.

Sudoyo, who also has an affiliation with the University of Indonesia, tells *Science* she will leave Eijkman and try to find a place where she can continue her research with her current co-workers. "What's most important for me is my team—these passionate and committed bright minds," she says. Sudoyo has several international grants; she hopes her funders will understand her dilemma and let her keep them. "This is force majeure," she says.

Protests have also erupted among technicians on the *Baruna Jaya*, a research vessel operated by the Agency for the Assessment and Application of Technology, which BRIN has also taken over. Its crew learned that 50 would have to leave after returning from Sumba island to implement an early warning system for tsunamis. "Before we departed, they told us that we could work for another 6 months. But now, they're forcing us to leave," says Andhika, a technician who has worked on the ship for 7 years. (Like many Indonesians, he goes by a single name.)

Scientists within BRIN, meanwhile, have complained that the new bureaucracy is hard to navigate. To set up an expedition, for example, they have to apply separately for travel and materials, a slow and cumbersome process, says Ibnu Maryanto, a biologist at BRIN. At the Herbarium Bogoriense, which has one of the world's largest collections of dried plants, researchers have had trouble getting access to samples after BRIN took over LIPI's botanical research department, Maryanto says. "I am afraid in the future, scientists will be reluctant to submit collections because of the red tape," he says.

Yet it's too early to judge BRIN, others say. "Every reorganization takes time," says Muhandis Shiddiq, a physicist formerly at LIPI and now at BRIN. Shiddiq applauds Handoko, who worked in Japan for many years, for his efforts to lure Indonesian scientists working abroad. "I hope this policy works," Shiddiq says. ∎

Dyna Rochmyaningsih is a science journalist in Palembang, Indonesia.

Omicron leads to fresh wave of meeting cancellations

One biology society goes ahead with in-person conference

By **Daniel Clery** *and* **Elizabeth Pennisi**

I
n a near rerun of early 2020, a handful of scientific societies are once again canceling their in-person meetings as the contagious Omicron coronavirus variant sweeps across the globe. Although many societies had planned hybrid meetings that could easily shed their in-person components, some were caught off-guard and forced to cancel altogether, whereas others doggedly plowed ahead with in-person components, taking extra safety precautions.

AAAS, the largest general science society in the world and publisher of *Science*, said last week it would cancel the in-person component of its annual meeting, set to take place next month in Philadelphia, and proceed with an online segment. AAAS CEO Sudip Parikh says the decision was meant to protect the health of staff, members, and attendees, and also to avoid contributing to the virus' spread. "As a scientific organization, it would be hard to justify," he says.

The American Meteorological Society announced that it, too, was canceling the in-person component of its annual meeting in Houston later this month. The Joint Mathematics Meetings said it would cancel a wholly in-person event set to begin last week in Seattle and replace it with a virtual meeting in April.

Similarly, the American Astronomical Society (AAS) canceled its annual meeting in Salt Lake City, upsetting plans for about 2200 people who had paid to attend a largely in-person event set for this week. AAS could not shift to a full virtual meeting in time, so the whole thing was abandoned. The society says its typically smaller summer meeting will be recast as the main event of the year. "We'll take a big financial hit—we're still unwinding that— but it's still the right decision, despite the aftermath," says AAS Executive Officer Kevin Marvel.

Some astronomers were frustrated and thought AAS should have been more prepared. "I was really just disappointed at the lack of contingency planning," says astronomer Elisabeth Mills of the University of Kansas, Lawrence, who lost grant money from nonrefundable hotel bills. "Having a robust virtual hybrid plan is going to be necessary for the next few years."

In contrast, the Society of Integrative and Comparative Biology (SICB) went ahead with its in-person meeting last week in the Phoenix convention center, which has a top-of-the-line ventilation system. Even as Omicron surged, in-person attendance numbers only dropped from 1500 to 1000. As added precautions, SICB moved audience seating farther from speakers in meeting rooms, and placed air filtration systems near the podium.

> ## "Having a robust virtual hybrid plan is going to be necessary for the next few years."
>
> **Elisabeth Mills**,
> University of Kansas, Lawrence

SICB also required attendees to show they had been vaccinated and wear masks. They placed posters farther apart and shifted social gatherings outdoors— not so hard to do in Phoenix in January. Some sessions were spiked or became virtual when speakers tested positive before the meeting. But the organizers worked hard—on New Year's Eve until 11 p.m.—to reorganize the program to avoid gaps. "It was a wild ride, and it was tough," says Jack Socha, a comparative biomechanist at the Virginia Polytechnic Institute and State University who oversaw setting up the program.

Most attendees seemed to be content with the experience. "I'm not so good at being remote," says Todd Oakley, an evolutionary biologist at the University of California, Santa Barbara. "And I like it that the meeting is a little less crowded." The true test will be in the numbers of COVID-19 cases among attendees in the coming weeks.

Parikh says he can't imagine future AAAS annual meetings without a virtual component. "There's too much to be gained," he says, in terms of increased participation and reduced travel time, costs, and carbon footprints. But he doesn't belittle the value of meeting friends and collaborators face to face. "We can do online for one more year," he says. "We'll be back in 2023." ∎

Study of soldiers implicates common virus as MS trigger

Vaccines under development against Epstein-Barr virus might prevent MS

By **Jocelyn Kaiser**

One hundred and fifty years after a French neurologist first recognized a case of multiple sclerosis (MS) in a young woman with an unusual tremor, the cause of this devastating disease remains elusive. Now, a study that combed data from regular blood tests of 10 million U.S. soldiers has found the strongest evidence yet that infection with a common virus, Epstein-Barr virus (EBV), dramatically increases a person's chances of developing the rare disease.

The work leaves many questions, such as why MS only affects about one in 1000 people even though nearly everyone will contract EBV in their lifetime. Still, "It provides probably the best evidence that can currently be obtained for a major pathogenic role of EBV in MS," says neurologist Hans Lassmann of the Medical University of Vienna, who was not involved in the study.

The study authors hope it will spur the development of a vaccine against EBV. The virus has been linked to several cancers and causes mononucleosis, and early vaccine testing is underway. Researchers then want to test whether vaccinating young people against EBV prevents MS.

MS develops when immune cells go awry and attack the myelin sheaths that insulate nerve fibers in the spinal cord and brain. The result is vision problems, pain, weakness, and numbness that can come and go, but worsen over time. Infusions of antibodies that deplete B cells, a type of white blood cell, can curb relapses. But the disease has no cure.

A combination of genetics—the disease often runs in families—and environmental triggers such as viruses is the likely cause. EBV, a herpesvirus that infects most people by adolescence and then lies latent in B cells throughout life, has long been a prime suspect. People who have had mono are at higher risk for MS. But although 99% of MS patients have had an EBV infection, 95% of those without MS have, too, making it difficult to pin down the virus' effects.

Ideally, researchers would track a group of young people who haven't yet been infected by EBV to see whether those who contract the infection are more likely to develop MS than those who don't. A team led by physician and epidemiologist Alberto Ascherio of the Harvard T.H. Chan School of Public Health found a clever way to do that. They probed a medical records database of 10 million active duty U.S. military personnel who enlisted between 1993 and 2003 and gave a blood sample every other year for HIV testing.

Eventually, 955 soldiers developed MS. Of the 801 with sufficient blood samples, 35 were negative for EBV in their first blood test; all but one became EBV positive during the study before developing MS on average 5 years later. By comparison, only half of 107 MS-free study participants used as controls became EBV positive during the same period, the researchers report this week in *Science*. That means an EBV infection multiplies a person's risk of MS 32-fold, comparable to the increase in risk of getting lung cancer from heavy smoking, Ascherio says.

None of the other common viruses Ascherio and his team tested for showed an effect. To bolster their case, they showed that people who eventually developed MS had a rise in levels of a protein linked to neural degradation after their EBV infection. Ascherio believes the study clinches the case. "How do you explain the fact that you don't get MS unless you get EBV? There is no other alternative explanation," he says.

Others are cautious. The new evidence is "very exciting," but "it's still an association," says Jeffrey Cohen, a virologist at the National Institute of Allergy and Infectious Diseases. And the study doesn't explain why most people who get EBV don't develop MS,

A new study links multiple sclerosis with Epstein-Barr virus, shown emerging from a B cell.

says neurologist Emmanuelle Waubant of the University of California, San Francisco. "Clearly other fuses have to be lit for the trigger to result in the disease," says Stanford University neuroimmunologist Lawrence Steinman, who co-authored a Perspective on the paper (DOI: 10.1126/Science.ABM7930).

Also dissatisfying is the lack of a known mechanism for how EBV might cause the immune attack. Some researchers suspect EBV transforms B cells so they become pathogenic; others, including Steinman, suggest an EBV protein resembles a neural protein and teaches the immune system to attack nerves.

An EBV vaccine could help researchers prove the virus has a causal role by vaccinating a large cohort of young people at high risk for MS because of family history. Experimental evidence that a vaccine prevents cases would "tick the final box," says neurologist Gavin Giovannoni of Queen Mary University of London, who is working with the MS patient community to design such a study.

Several years ago, GlaxoSmithKline developed a vaccine based on an EBV envelope protein but abandoned it after a trial showed it reduced the incidence of mono but didn't prevent EBV infections. Two new candidate vaccines now in early clinical trials could be more potent. One developed by Cohen displays the same EBV protein on nanoparticles. Another from Moderna contains messenger RNA that instructs cells to make four different EBV proteins.

Vaccine expert Larry Corey of the Fred Hutchinson Cancer Research Center cautions that despite the potential public health benefits, there's no guarantee a company will take an EBV vaccine through licensing. Still, the new evidence firming up the role of EBV in MS "should make the risk benefit of that investment much greater," he says. ∎
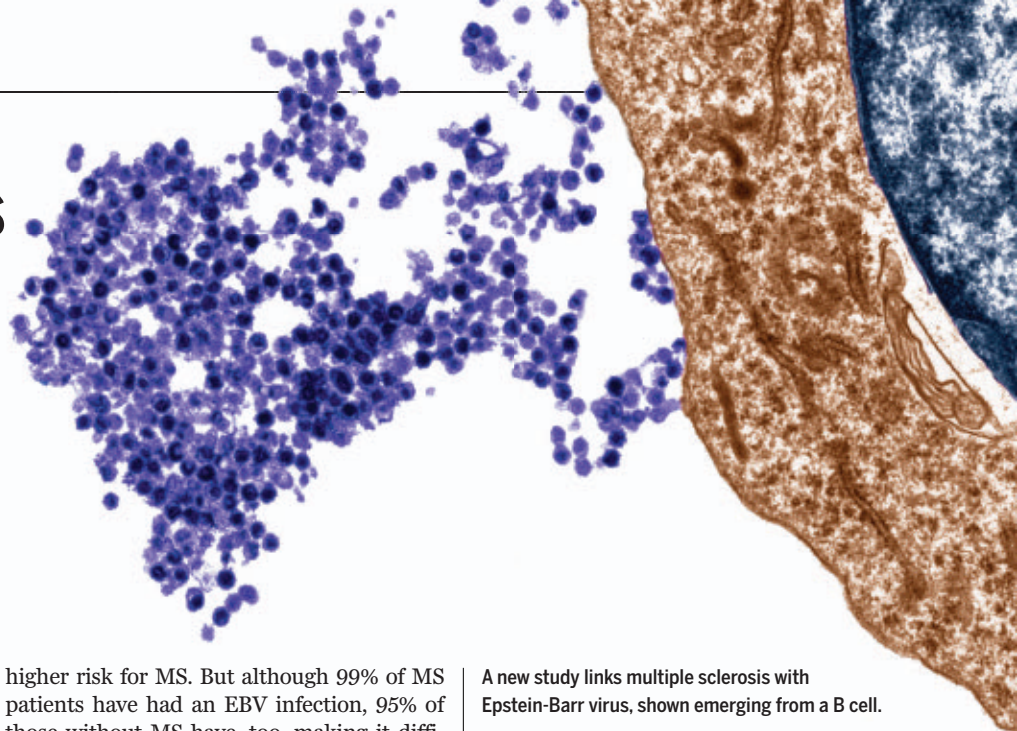
CLONING GOES WILD

A ferret named Elizabeth Ann
could become the first cloned mammal
to help save an endangered species

By **Rachel Fritts**

Last month, at a conservation center near Fort Collins, Colorado, staffers held an unusual birthday party, complete with a two-tiered cake made of prairie dog and mouse carcasses, minced meat, and kibble. The recipient of the macabre cake was a small, weasellike animal named Elizabeth Ann. She is the world's first cloned black-footed ferret, one of North America's most endangered species, and her first birthday was a major milestone: She is one of the first clones of an endangered species to reach sexual maturity.

Now, Elizabeth Ann—cloned from the cells of a female ferret that died 35 years ago—is poised to make history again. This spring, if all goes as planned, Elizabeth Ann will mate with a carefully selected bachelor in an effort to introduce greater genetic diversity into wild ferret colonies, which are threatened by inbreeding. If she gives birth to healthy kits, it will mark the first time conservation biologists have been able to integrate cloning into an effort to save a species from extinction.

Success could boost nascent efforts to clone other endangered mammals, including rhinos, and help establish the technology as a useful restoration tool. Failure could reinforce long-standing skepticism about the usefulness of cloning in conservation; some researchers see it as too expensive, ethically fraught, and of limited use. They also fear it could distract funders from efforts to address broader issues such as habitat destruction.

"Everything about Elizabeth Ann is much bigger than the science behind it, and it's much bigger than helping the ferrets," says Ben Novak, black-footed ferret project lead for Revive & Restore, a nonprofit founded in 2012 to explore how biotechnology might aid endangered and extinct species. "It's about whether biotechnology can become a part of mainstream conservation."

THE BLACK-FOOTED FERRET (*Mustela nigripes*) is a slender, half-meter-long predator with an ornery streak. It once inhabited huge swaths of the Great Plains, occupying grassland burrows dug by its favorite prey: prairie dogs. By the 1970s, however, the widespread destruction of prairie dog colonies by ranchers, farmers, and others had caused ferret populations to crash. In 1973, the ferret became one of the first species to be designated for protection under the new U.S. Endangered Species Act.

By the late 1970s, the last known ferret colony had disappeared and some biologists believed the species was extinct. But in late 1981, the Wyoming Game and Fish Department received an unexpected call.

One morning a rancher named John Hogg had gone out to investigate the origins of some strange noises heard during the night. He suspected the family dog, Shep, had picked a fight with some varmint. Dead on the ground was a strange, tube-shaped animal that a local taxidermist identified as a black-footed ferret. After wildlife biologists descended on the area around the Hogg ranch, they were overjoyed to discover a good-size colony of more than 100 ferrets.

But within a few years, that colony, too, was in trouble, reduced to just a few dozen animals. In 1985, officials made the difficult decision to round up any ferrets they could find, in hopes of starting a captive breed-

ing program. They captured 18 individuals, but just seven survived to breed, putting the species at risk of inbreeding, which can erode reproductive fitness.

ELIZABETH ANN owes her existence to a chance meeting at a banquet in Montana. In 1987, biologist Oliver Ryder was on the lookout for animal cells to freeze. As a young scientist in the 1970s, he had joined a new effort, known as the San Diego Frozen Zoo, that aimed to preserve genetic material from a wide array of endangered mammals, deep-frozen in liquid nitrogen. At a conservation conference, Ryder struck up a conversation with a Wyoming Game and Fish veterinarian named Tom Thorne, who told him about the black-footed ferrets' plight. A couple of years earlier, Thorne had sent a handful of black-footed ferret tissue samples to the Frozen Zoo, but scientists had managed to preserve a cell line from just one: a male labeled Studbook #2. Ryder suggested Thorne send more. Seven months later, he received skin cells from a single female named Willa, which the zoo also successfully banked.

As those two cell lines sat in a freezer, the ferret breeding program—which has been led by the U.S. Fish and Wildlife Service (FWS) since 1996—began to take off. To date, it has produced some 10,000 ferrets, many of which have been released into the wild. The program has become a prominent success story, but the ferret's survival is still far from a sure thing.

One threat is sylvatic plague, a deadly bacterial disease introduced from Asia. Inbreeding has also begun to take a toll on the population, all descendants of just seven animals. Genetic studies, Novak says, show "they're all superrelated. … The mean kinship falls between that of a sibling and a first cousin." As a result, they have accumulated some potentially damaging mutations, says genomic researcher Klaus-Peter Koepfli with the Smithsonian Mason School of Conservation. Some animals are born with kinked tails and deformed sternums.

To shore up the ferrets' genetic diversity, researchers have turned ferret breeding into an exacting science. A computer program helps them assign a desirability rating to each possible pairing, based on the ani-

Biologists hope Elizabeth Ann, a cloned black-footed ferret (above), will soon contribute to efforts to breed the endangered species and restore it to grasslands.

mals' ancestry. Breeders have also relied on artificial insemination with semen taken in the 1990s from two standout males, known as Scarface and Rocky. But semen is a finite resource, says Pete Gober of FWS, who coordinates the recovery programs.

So, in 2013, program managers turned to Revive & Restore to see whether they could bolster the population's diversity by transforming the Frozen Zoo's small vials of preserved cells into living, breathing ferrets. "We wanted to increase and maintain as much genetic diversity as we could from what little amount we had to begin with," Gober says.

THE IDEA WAS to use a technology that didn't exist when the zoo had stored the samples: somatic cell nuclear transfer. In this cloning technology, technicians replace the nucleus of an egg cell with a nucleus taken from a body cell. A jolt of electricity encourages the egg and nucleus to fuse and the cell to

multiply; the embryo is then transferred to a surrogate mother. In 1995, Dolly the sheep was the first animal to be born from cloning. Since then, cloning has become a routine tool for duplicating farm animals, pets including cats and dogs, and even prized racing camels.

Cloning for conservation has a far patchier history. When black-footed ferret breeders first approached Revive & Restore, for instance, researchers had created clones of just three endangered species, all hoofed animals: the gaur (*Bos gaurus*) and European mouflon (*Ovis aries musimon*) in 2001, and the banteng (*B. javanicus*) in 2003. In 2015, scientists cloned a fourth endangered species, a sheep known as the Esfahan mouflon (*O. gmelini isphahanica*). All died fairly young (the banteng lived longest, dying of injuries at age 7), and produced no offspring.

In part, cloning endangered animals has proved harder than duplicating livestock or pets because breeding and husbandry practices are less developed in these species. Conservation programs also have fewer resources than commercial enterprises, so they're less likely to try again after an unsuccessful attempt.

Cloning endangered species faces unique ethical questions, as well. One is whether the clone, which can hold trace DNA from its surrogate mother, is actually the same as the species that researchers are trying to save. For example, black-footed ferret clones are created using eggs from domestic ferrets, meaning they carry that species' mitochondrial DNA, which is left in the egg after its nucleus is extracted.

Some conservationists have other concerns. They worry the ability to clone a rare species might undermine support for efforts to protect habitat and keep species alive in the wild. And cloning can be expensive, potentially diverting funds from other conservation activities.

"I think cloning certainly has a future for endangered species, but there are some problems," says Barbara Durrant, director of reproductive sciences with the San Diego Zoo Wildlife Alliance. For each species, "We will need to be very careful to do the basic research," she adds.

A few years ago, after extensive technical and ethical reviews, federal regulators decided the potential benefits of cloning the ferrets outweighed the risks. Then, with permits in hand, Revive & Restore teamed up with firms including a pet cloning company called ViaGen Pets and a commercial ferret breeder to develop a plan that cost about $40,000 to execute. It called for creating embryos with DNA taken from Willa, the female black-footed ferret that had died in 1988.

Deepening the gene pool

By cloning a black-footed ferret that died in 1988, biologists hope to add greater genetic diversity to the existing population of this endangered species, which relies heavily on captive breeding for survival.

1985–87

In Wyoming, biologists capture 18 of the last known wild ferrets to start a captive breeding program (left). They freeze cells from one male and one female (right).

Colored bands represent genetic variation.

Biologists hope to clone the male in the future.

Cloning efforts have focused on the female, named Willa.

Only seven bred successfully.

Breeding has produced more than 10,000 ferrets. All are descendants of those seven. Genetic variation declined by 55%.

2020

Willa's DNA was used to produce a clone, Elizabeth Ann.

2022

Biologists will try to breed Elizabeth Ann with a captive male. Offspring could add Willa's DNA to the population, increasing genetic diversity and improving the odds of survival.

Even ViaGen's lead scientist, Shawn Walker, wasn't sure whether it would work. "We knew we had all the boxes checked," he says. "But until you hear that heartbeat, you're always a little bit skeptical, because everything was so new."

In late 2020, the team implanted Willa-based embryos into three domestic ferrets and shipped them to the National Black-footed Ferret Conservation Center in Colorado, where about two-thirds of the nation's captive population lives. To their relief, one ferret gave birth to Elizabeth Ann on 10 December 2020. "It was very exciting," says Robyn Bortner, captive breeding manager at the facility, who was in the room when the clone was born.

But success still wasn't a sure thing. The other two pregnancies failed, and Elizabeth Ann had a stillborn litter mate. And when keepers placed Elizabeth Ann with a second surrogate mother and domestic siblings, she often ended up in a dangerous spot: at the bottom of the pile of kits. "It was touch and go the first couple of days. … We kept an incredibly close eye on her," Bortner recalls. But once Elizabeth Ann's eyes opened about 1 month later, her black-footed ferret feistiness kicked in. Ever since, Bortner says, "She's been healthy and everything you would hope."

NOW FULLY GROWN, Elizabeth Ann looks and sounds like any other black-footed ferret. She scarfs down meat and stalks and kills live hamsters. (The keepers aren't willing to risk giving her a live prairie dog, which weighs more than she does.) She chatters angrily at caretakers who get too close. She loves to attack and shred paper bags. And apart from her mitochondrial DNA, most of which comes from her domestic mother,

genetic analysis shows she is 100% a black-footed ferret.

This spring her creators hope to mate Elizabeth Ann with a captive male. Any offspring will still have Elizabeth Ann's mitochondrial DNA, with traces of domestic ferret. To remove those traces, any male offspring will be paired with captive females, producing kits that no longer carry the domestic female's mitochondrial DNA.

Successfully adding Willa's genes to the black-footed ferret gene pool via Elizabeth Ann would likely "pack this huge biodiversity punch," Novak says. Genomic analysis has found Willa's DNA has 10 times more unique alleles than DNA from any captive-bred ferret. That means her chromosomes will "introduce a whole new combination [of traits] and higher level of genetic variants," Koepfli says. That should slow the ferrets' trajectory of reproductive decline.

To introduce even more genetic variation into captive-bred ferrets, ViaGen aims to create company for Elizabeth Ann. Scientists there are fine-tuning the cloning procedure to make it more efficient and will try to create the next batch of cloned ferrets in spring 2023, to align with next year's breeding season.

THE SUCCESSFUL USE of cloning in ferret conservation is likely to attract attention—and perhaps funding—for similar efforts in other endangered species. But replicating it won't be easy. In part, that's because species that might benefit from cloning have to meet numerous criteria. The best candidates, for example, have both banked genetic material at the ready and a less endangered close relative that can act as a surrogate. It also helps to have funding and captive breeding infrastructure in place. Few programs can meet these prerequisites—yet.

Another obstacle is that, although the basic cloning process is the same for all mammals, the technology has worked better in some species than others, and "no one knows why," Durrant says. To increase the odds of success, researchers often must develop a unique "recipe" that addresses an animal's reproductive quirks, says Samantha Wisely, a conservation geneticist at the University of Florida who works with the black-footed ferret program. "Reproductive technology is super–species specific."

Still, efforts to clone at least two other endangered species are underway. One is the Przewalski's horse (*Equus ferus przewalskii*), a stocky wild horse that once roamed across Europe and Asia. The species nearly went extinct in the mid–20th century, and all individuals alive today are descended from just 12 animals. Luckily, nearly 300 cell lines have been stashed at the

Frozen Zoo, and conservationists are now trying to inject some of that lost genetic diversity into the modern population.

In 2020, researchers created Kurt, the clone of a Przewalski's horse whose cells were frozen 40 years ago. Although he was born several months before Elizabeth Ann, he still has some growing up to do before he'll be ready to breed. Kurt could be joined by cloned siblings by spring of 2023.

The San Diego Zoo Wildlife Alliance is also in the preliminary stages of trying to clone the northern white rhino (*Ceratotherium simum cottoni*), the most endangered of several rhino subspecies. Just two living northern white rhinos remain, and

neither is capable of giving birth. As a first step, scientists are working on techniques to incorporate the northern white rhino genome into egg cells of a close relative, the southern white rhino (*C. simum simum*). And once the team hones its techniques, "We certainly hope we can apply them to the black rhino, the Sumatran rhino, and maybe the Javan rhino," Durrant says.

Some researchers are looking beyond cloning, to other genetic technologies that might help endangered species. Birds, for example, can't be cloned, but Revive & Restore recently formed a research consortium to develop a technique that could fill a similar role. It involves introducing primordial germ cells from an endangered species into an embryo of a surrogate species, such as a chicken. These germ cells then migrate to the chicken's gonads and become sex cells. So, a male domestic chicken could

produce the sperm of, say, an endangered prairie chicken.

CRISPR gene-editing tools could also play a role in conservation. Editing the genome of black-footed ferrets so that they could resist sylvatic plague, for example, could be a game changer. Koepfli and collaborators are now comparing the black-footed ferret genome with that of its domestic cousin, which plague does not affect, in hopes of identifying the genetic basis of resistance. But actually genetically modifying ferrets and then setting them loose in the wild would require extensive legal and ethical deliberations.

For now, Elizabeth Ann's caregivers are simply gearing up to pick her ideal first

At the San Diego Frozen Zoo, the cells of numerous endangered species are preserved in liquid nitrogen. Cloning a mammal was not yet feasible when the zoo banked cells from two black-footed ferrets in the 1980s.

mate. Every male ferret in the six breeding facilities scattered across the United States and Canada is getting intense scrutiny. In part, that's because Elizabeth Ann is likely to have just a few good breeding years, and annual litters average only three to five kits. Breeders have decided that, first and foremost, they need a proven gentleman—they can't risk an aggressive ferret hurting their only clone. Excellent genes matter, too. (That criterion could put a male related to Scarface or Rocky in the mix, Novak muses.)

This month, they will develop a short list and make their pick. If the best fit happens to live across the country, they will fly him to Colorado. They'll collect a semen sample, just in case they need it to artificially inseminate Elizabeth Ann. But they are hoping that once they put the two ferrets together, nature will simply take its course. ■

POLICY FORUM

CHEMICAL REGULATION

Legal obstacles to toxic chemical research

Legislative design impedes study of chemicals in the environment

By **Wendy E. Wagner**[1] *and* **Steve C. Gold**[2]

When the United States first regulated industrial chemicals in the 1970s, policy-makers made a threshold decision: Rather than a "front-end" approach requiring proactive risk assessment before licensing a chemical for use, Congress chose a reactive approach of regulating chemical risks as they became manifest (*1*). This choice puts a premium on "back-end" research on exposure pathways and hazards of new chemical contaminants in situ. Such research identifies contaminants in the environment, their source, where they go, and what harm they might cause (*2*). This knowledge is crucial for risk assessment and regulatory decisions. However, key features of the legislative design impede that essential research. Amid debate about how policy-makers should apply existing scientific findings, these science-hindering features are easy to overlook. We highlight three of them: insufficient availability of chemical standards, limited public access to information, and excessive fragmentation of information within and among government agencies.

A recent survey found that 22 chemical databases and inventories worldwide

When new contaminants are found in environmental samples like the one shown here, the law places obstacles in the way of further analysis.

of US chemical regulation—also found to some extent in other regulatory systems—serve to frustrate and delay rather than facilitate the needed back-end research to identify, analyze, and assess new chemicals in the environment. A case study involving a group of chloroperfluoroether carboxylates (Cl-PFECAs) detected in New Jersey (5) highlights these issues (see the box).

INSUFFICIENT AVAILABILITY OF CHEMICAL STANDARDS

Given the pace of chemical innovation and the absence of much identifying information from registries, back-end research depends on a range of nontargeted analytical techniques to determine the presence of emerging and new contaminants (2). Chemical reference standards are important to this work (1, 6). Without a chemical standard, researchers are precluded from definitively identifying a chemical found in the environment. Even if a chemical can be identified, measuring its concentration requires a standard for comparison. In cases of complex mixtures of chemicals, the need for chemical standards is even more crucial to provide benchmarks for the research (2). Without chemical standards, this first step to back-end field research is always imprecise, labor-intensive, and time-consuming [(1); see supplementary materials (SM)].

Despite the critical importance of analytical reference standards to researchers, US law does not require manufacturers to provide a chemical standard as a condition to marketing a chemical (1). Sometimes chemical reference standards do exist but are classified as trade secrets (1, 6).

If a manufacturer does not share a reference standard of its chemical, field researchers generally have two options. First, under ideal conditions, scientists may be able to determine the identity (but not the concentration) of an unknown chemical in a complex mixture. However, this approach takes a great deal of additional time and may be available only to a subset of researchers who have high-end laboratory equipment and expertise (1). Alternatively, for a fee, researchers may be able to obtain standards of some chemicals purchased directly from third-party companies that synthesize and sell certified reference standards.

But even when intermediate groups sell these reference standards, legal impediments may still arise. In 2020, a manufacturer sent a cease-and-desist letter to a company selling a reference standard of its chemical, arguing that marketing the stan-

dard infringed the manufacturer's patent (6). Given the relatively small profits from that particular reference standard, the standards company, which at the time was the sole provider of the standard, chose simply to stop selling the standard rather than agree to the manufacturer's proposed licensing agreement (6, 7). The mere threat of legal action, whether based on a valid claim or not, can discourage companies as well as individual chemists from creating or selling certain chemical reference standards. And a chemical manufacturer can always argue that independently created standards are imprecise, unreliable, or even illegal (1). Under this legal design, manufacturers can effectively set "the rules for which chemicals environmental researchers can and can't measure" (7).

LIMITED PUBLIC ACCESS TO INFORMATION

Once a chemical in the environment is identified, field researchers will attempt to learn how, where, when, and why it is produced and released. At this second step, the field researcher can encounter another legal obstacle: limited public access to crucial information, even if the information is somewhere in a government agency's files. The law keeps some submitted information secret and allows other information to be submitted in a form too imprecise to be truly useful.

The best-known impediment to accessing information arises when a manufacturer classifies some or most of a submission as a protected trade secret or confidential business information (CBI). Under the Toxic Substances Control Act (TSCA), CBI protection can even apply to the identity of the chemical itself, precluding researchers from learning whether the chemical has been registered with the US Environmental Protection Agency (EPA) (1). As of August 2021, the identities of nearly 20% of the active chemicals on the EPA's TSCA inventory—more than 8200 chemicals—were classified as CBI (see SM). And even for publicly listed chemicals, a great deal of information can be classified, including the name of the manufacturer, the location of its operations, chemical trade names, and process information (see SM).

Not all information on a chemical may be protected as a trade secret; the TSCA provides that health and safety information on a chemical cannot be classified as CBI because of its public import. Notwithstanding this limitation, however, the toxicity data may be effectively shielded from public view if the chemical identity itself is deemed CBI. In the case of Cl-PFECAs, toxicology studies in the EPA's files came to light only after Washington *et al.* (5) publicly described the chemicals and a journalist filed a Freedom

included more than 350,000 registered chemicals or chemical mixtures, the vast majority of which are or were in commerce (3). About one-third are not fully identified, and about one-fifth—70,000 chemicals or mixtures—were registered in just the past decade (3). The number of chemicals continues to grow exponentially (4).

With ever more new chemicals and their degradation products possibly making their way into the environment, identifying emerging contaminants is both increasingly urgent and challenging (2). However, despite this urgency, several key features

¹School of Law, University of Texas at Austin, Austin, TX, USA. ²Rutgers Law School, Newark, NJ, USA. Email: wwagner@law.utexas.edu

of Information Act request that triggered an internal EPA review of the manufacturer's existing CBI claims.

In designing the TSCA in 1976, Congress sought to encourage chemical innovation in part by providing generous legal protections for information that, according to manufacturers, must be kept secret to maintain a competitive edge (*1*, *8*). Under this law, manufacturers face no consequences for overclaiming trade-secret protection. Indeed, historically, manufacturers were not even required to substantiate or justify their claims in advance (*9*). By contrast, manufacturers irrevocably lose the legal right to confidential treatment of any information submitted without a trade-secret claim (*9*). Faced with these asymmetrical incentives, many manufacturers historically overclassified information as trade secret–protected (*9*). Even after the TSCA's 2016 amendments, which required more substantiation of some CBI claims, it is up to the EPA to assess the legitimacy of each CBI claim and to take appropriate action if the agency believes a claim is unwarranted. Generally speaking, substantiation requires an explanation of how disclosure could harm a submitter's competitive position and of the steps a submitter has taken to keep the information secret (see SM). For CBI claims that the EPA has reviewed since 2016, more than 32% were considered unsubstantiated and hence rejected (*10*).

From the standpoint of non-EPA scientists researching the chemical fate of new chemicals in the field, these trade-secret protections can serve as a barrier to important data and related information about the chemical (*11*). Even EPA scientists are not able to access CBI information unless they have been legally cleared to view CBI materials [(*9*); see SM]. Yet clearance to view classified information is a double-edged sword. The law prohibits a researcher from disclosing the confidential information in subsequent publications, even indirectly. So, if a manufacturer can make a case that a government employee somehow disclosed CBI, or disclosed information that could have been uncovered only with foreknowledge of CBI, then that employee may be subject to civil lawsuits or criminal prosecution and potential imprisonment (*8*, *9*). As Richter *et al.* report, "The penalties for violating CBI policy transform basic types of scientific inquiry into potentially non-normative, criminal behavior" [(*1*), p. 12].

A second set of impediments to scientists' efforts to access chemical information in government files arises because regulatory regimes tolerate substantial ambiguities in the identification of some subsets of chemical substances (*3*). These imprecisions tend to arise early in the regulatory process, when manufacturers first register a substance with

an agency. For example, manufacturers may identify substances of unknown or variable composition, complex reaction products, or biological materials (UVCBs) with a generic name that does not reveal such substances' composition (*3*). Scientists may similarly search in vain for information on a substance that they have identified if the manufacturer registered and submitted data on a slightly different structure or composition and thus a different Chemical Abstracts Service (CAS) number: an anion versus a salt, for example (as with Cl-PFECAs), or salts with different cations. Addressing these additional information gaps is "an important next step to advance the current chemical registration and assessment schemes" [(*3*), p. 2580)].

EXCESSIVE FRAGMENTATION OF INFORMATION WITHIN AND AMONG GOVERNMENT AGENCIES

Scientists doing "back-end" environmental research would, ideally, have full access to all available information on toxicity

and environmental fate and transport (*2*). But even when this essential information exists (which is by no means assured, because manufacturers generally have no legal duty to produce it), the information is badly fragmented and siloed (*12*). Different agencies at different levels of government, and even different offices within the EPA, each receive different bits of information as a result of a series of disconnected legal requirements. Obtaining a comprehensive picture of a chemical thus requires effective coordination within and among agencies and, in many cases, considerable ingenuity of researchers.

For example, for chemical registration purposes, the TSCA requires manufacturers to submit toxicity information to the federal EPA, but that information does not necessarily reach state governments responsible for public safety in the places where registered chemicals are used. Indeed, information submitted to the EPA's chemical regulation office may

---

Searching for answers, coming up empty

In May 2020, a team of researchers (*5*) from the US Environmental Protection Agency (EPA) and the New Jersey Department of Environmental Protection identified a group of chloroperfluoroether carboxylates (Cl-PFECAs) in soil samples from parts of New Jersey. These particular per- and polyfluoroalkyl substances (PFAS) were unknown to the EPA members of the research team, who used nontargeted methods to perform analytical work. Using semiquantitative techniques in the absence of an analytical standard, they plotted contours that showed Cl-PFECA concentrations descending with increasing distance from a putative source facility. But the EPA chemists could learn very little else about what they had found, even though they were able to find a Chemical Abstracts Service (CAS) number associated with these Cl-PFECAs.

We attempted to merge Washington *et al.*'s (*5*) discovery with existing regulatory information. Our searches of US government databases for these Cl-PFECAs by CAS number came up empty, although the substance had been approved by the European Food Safety Authority (EFSA) for use in manufacturing nonstick coatings and is subject to the European Union (EU)'s chemical labeling regulation (*5*). We ultimately learned from Freedom of Information Act requests filed by others and by us that the anions that Washington *et al.* described are not registered in the United States, although at least three related Cl-PFECA salts and esters are known to the EPA—only one of which could be found in the EPA's public database of regulated chemicals [see supplementary materials (SM)]. We also learned that two of the substances have been linked to evidence of potential bioaccumulation in the blood of workers documented by the manufacturer since 2011 (see SM).

To our knowledge, but for the filing of formal information requests, all this information would still be treated confidentially and concealed from the public. But even after formal information requests were fulfilled, many mysteries remained in the redacted information released by state and federal agencies (see SM). For example, researchers still cannot learn the who, what, where, why, and when of the use or production of these compounds. Even the toxicological studies are partly redacted and hence incomplete.

This situation is not specific to the EPA or to US law. The three toxicology studies that we received from the EU's EFSA, about 3 months after requesting them, were similarly redacted (see SM). And under the EU chemicals regulation Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH), manufacturers that register chemicals are required to submit only "robust summaries" of toxicity studies, not the studies themselves (*3*). Even if the full, unredacted studies were ultimately available, however, it is unclear whether scientists interested in replicating or critically reviewing the research would be able to do so (*15*).

---

not even reach other offices within the EPA, such as the air and water pollution program offices. Conversely, if a chemical in commerce is classified as hazardous, a different law requires facilities to inform state and local agencies, but not the federal EPA, if the chemical is present at the facility above threshold amounts (see SM). Other state and local legal authorities may allow some agencies to obtain information on an ad hoc basis, as New Jersey did with Cl-PFECAs (see SM). The result is a patchwork in which, for a given chemical, individual EPA offices and states may get different toxicity information; states but not the EPA may get information about where the chemical is used (but states might not learn the chemical's identity); and only if the chemical has already officially been deemed toxic do both states and the EPA get information about releases above a threshold amount.

Trade-secret claims further complicate the statutes' built-in information fragmentation. The 2016 amendments to the TSCA clarified that the EPA may share CBI with state, Tribal, and local authorities that need the information for implementation or enforcement of a law. To our knowledge, however, no state agency has finalized an intergovernmental agreement with the EPA that would enable sharing CBI as a matter of course. On the federal level, the EPA and the Occupational Safety and Health Administration entered into an agreement to share confidential business information only in January 2021 (see SM). We could not locate any international sharing agreements and therefore expect that access to CBI-protected information is likely fragmented internationally as well.

The statutory design thus requires entrepreneurial engagement by researchers, well beyond the usages of ordinary scientific collaboration, to ensure that all available information is accessible to them. Nongovernment investigators, in particular, may not understand that federal and state agencies hold different data or might not know which state or EPA office to ask for the data, if the presence of a chemical at a specific facility is a protected trade secret. The extra coordination burdens may inhibit or limit some research projects. Even if researchers are willing to make the effort, the problem persists: If the chemical identity information is classified, even agency staff may not know to look for it.

REFORM

Explanations for why these legal impediments have arisen are incomplete. Based on the current literature, the leading hypothesis is that the manufacturing community was very involved in the development of both the original 1976 TSCA statute and the 2016 amendments and carefully secured some benefits in the legislation that enabled industry to retain control over certain key information (1).

How can these legal impediments to back-end research be overcome? More research, and probably legislative action, is needed to craft comprehensive solutions. We propose some interim measures that the EPA might take in the meantime.

For the lack of analytical reference standards, the solution seems obvious. The EPA should require manufacturers to provide a standard, to be made generally available to the public, for each registered chemical—perhaps as a mandatory condition of registration but certainly, at least, upon request (1, 7). A mandatory requirement for reference standards has long been in place for pesticides found in residues on foodstuffs, which, to our knowledge, has neither imposed expensive burdens on the manufacturers nor proved difficult to administer (13).

To begin addressing the problem of limited public access to information on file with a government, the EPA could take several stopgap measures using its legal authority under the TSCA (8). First, the EPA should help educate researchers by providing more accessible information about the extent of CBI claims and the sources of imprecision in chemical identification. Second, the EPA should interpret the TSCA to allow the automatic release of toxicological information on all chemicals, including those with CBI-protected identities (see SM). Third, the EPA should provide needed information to researchers by making better use of section 14(d) of the TSCA, which allows disclosure of CBI to protect against "an unreasonable risk of injury to health or the environment" (see SM). Finally, the EPA should develop more systematic and consistent chemical identification methods (3). All of these measures would help, but the legal treatment of CBI claims about chemicals, and the challenges of chemical identification, merits systemic legislative reform (8, 9).

Finally, to address the problem of fragmented information, the EPA should use its own authorities and collaborate with states. An example of at least a partial solution is already in the regulatory pipeline. Acting pursuant to a new and specific statutory directive, the EPA has proposed to require manufacturers of more than 1000 per- and polyfluoroalkyl substances (PFAS) to report to the EPA both past (back to 2011) and future information on the production, use, disposal, releases, exposures, and toxicity of these substances (14). This type of broader reporting requirement could simply be extended to manufacturers of all chemicals, not limited to PFAS, produced in quantities above a low threshold amount (see SM). States (after an initial request) could be automatically notified of all toxicological information obtained by the EPA. Conversely, states could provide the EPA with facility-level data that the states receive (see SM). These steps would not ensure that agencies take a comprehensive view of chemical regulation, but they would help alleviate the information fragmentation problem.

These are preliminary suggestions for measures to provide researchers with greater access to information that is essential for back-end research while still protecting legitimate confidentiality interests. We have not comprehensively analyzed approaches taken in other countries, but our research suggests that the impediments we discuss here are not wholly specific to the United States. Specific reforms must, of course, be tailored to each government's structure and procedures. Regardless of such details, the most essential next step is to engage in an interdisciplinary investigation of impediments to research on chemicals in the United States and abroad. ∎

REFERENCES AND NOTES

1. L. Richter, A. Cordner, P. Brown, *Sociol. Perspect.* **64**, 631 (2020).
2. L. Chibwe, I. A. Titaley, E. Hoh, S. L. M. Simonich, *Environ. Sci. Technol. Lett.* **4**, 32 (2017).
3. Z. Wang, G. W. Walker, D. C. G. Muir, K. Nagatani-Yoshida, *Environ. Sci. Technol.* **54**, 2575 (2020).
4. E. J. Llanos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12660 (2019).
5. J. W. Washington *et al.*, *Science* **368**, 1103 (2020).
6. C. Hogue, C. Bettenhausen, *Chem. Eng. News* **99**, 14 (2021).
7. L. Stringer, "Solvay refutes NGO claims it is impeding analysis of its PFAS," Chemical Watch (3 March 2021); https://chemicalwatch.com/223735/solvay-refutes-ngo-claims-it-is-impeding-analysis-of-its-pfas.
8. M. L. Lyndon, in *The Law and Theory of Trade Secrecy: A Handbook of Contemporary Research*, R. C. Dreyfuss *et al.*, Eds. (Edward Elgar, 2011), p. 442.
9. W. Wagner, D. Michaels, *Am. J. Law Med.* **30**, 119 (2004).
10. EPA, TSCA CBI review statistics; www.epa.gov/tsca-cbi/statistics-tsca-cbi-review-program#stats.
11. S. Coffin, H. Wyer, J. C. Leapman, *PLOS Biol.* **19**, e3000932 (2021).
12. R. E. Lundgren, A. H. McMakin, *Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks* (Wiley, ed. 6, 2018).
13. EPA, "Residue testing guidelines" (EPA 712-F-08-001, EPA, 2009).
14. EPA, "Toxic Substances Control Act reporting and recordkeeping requirements for perfluoroalkyl and polyfluoroalkyl substances" (Proposed Rule EPA-HQ-OPPT-2020-0549-0001, EPA, 2021); www.regulations.gov/document/EPA-HQ-OPPT-2020-0549-0001.
15. European Chemicals Agency (ECHA), "Dissemination and confidentiality under the REACH Regulation" (ECHA, 2021).

FUNDAMENTAL PHYSICS

Quantum probe of space-time curvature

An atom interferometer measures the quantum phase due to gravitational time dilation

By **Albert Roura**

In Einstein's theory of general relativity, gravity is a manifestation of space-time curvature. As predicted by general relativity and confirmed by numerous measurements, clocks moving at different velocities or located in different regions of a gravitational field tick at different rates (*1*), a phenomenon known as relativistic time dilation. Under appropriate conditions, time dilation can affect the oscillation phase of quantum waves and give rise to a measurable effect in interference experiments. On page 226 of this issue, Overstreet *et al.* (*2*) present an atom interferometry experiment in which this effect has been measured for gravitational time dilation. In addition to the importance of the results for fundamental physics, the methods used can lead to more accurate measurements of Newton's gravitational constant, which parametrizes the strength of the gravitational interaction and is by far the least accurately known of all fundamental constants (*3*).

In quantum mechanics, microscopic particles can behave as waves, and each particle is characterized by a "wave packet." Forces modify a wave packet's propagation in the same way they would alter a particle's trajectory in classical mechanics. However, uniform changes to the potential energy can modify the oscillation phase of the wave packet without affecting its trajectory—a phenomenon with no classical counterpart. As early as the 1950s, Aharonov and Bohm (*4*) conceived an interferometry experiment with charged particles to observe this quantum effect. Since then, several versions of the experiment involving electromagnetic fields have been realized (*5*, *6*). By contrast, analogous measurements for the much weaker gravitational interaction had remained elusive and have only been possible thanks to extremely sensitive atom interferometers with arm separations of up to half a meter (*7*).

Atom interferometers rely on the wave nature of quantum particles and can serve as highly sensitive inertial sensors for both fundamental physics measurements and practical applications (*8*). In the atomic fountain setup used by Overstreet *et al.*, the atoms are launched vertically at the bottom of a 10-m vacuum tube and follow a free-fall trajectory (see the figure). Short laser pulses are applied at different times and act as light gratings that split, redirect, and recombine the atomic wave packets. Each atom is thus in a quantum superposition simultaneously following two different trajectories, sometimes referred to as the upper and the lower arm. Differences between the phase changes experienced by the wave packets as they evolve along the two arms can be read out from the interference signal.

There are two kinds of contributions to these phase changes. One corresponds to the propagation of the wave packets and is proportional to the proper time along each arm, which is the time that an ideal clock following the same trajectory would measure and includes relativistic time-dilation

Interferometry experiment in an atomic fountain

The atoms are launched vertically at the bottom of a 10-m vacuum tube and follow a free-fall trajectory. Laser pulses were applied at three different times to split, redirect, and recombine the atomic wave packets. The gravitational influence of the ring mass on the upper interferometer arm can be detected in the interference signal.

30 km/h

t = 0 s   *t* = 0.8 s   *t* = 1.6 s

effects. The other contribution is connected to the laser pulses. Every time a wave packet is diffracted by a laser pulse, it gets a momentum kick, but it also experiences a phase change that depends on its position with respect to the light-grating wavefronts.

About a decade ago, researchers proposed a hypothetical experiment for realizing a gravitational analog of the Aharonov-Bohm experiment (*9*). In this proposal, atoms in the two arms spend a sufficiently long time at two specific points where the net gravitational force from a pair of massive spherical shells vanishes. Yet, the different value of the gravitational potential at these two points leads to a measurable phase difference. Within the framework of general relativity, this phase difference corresponds to the proper-time difference between the two interferometer arms due to gravitational time dilation. Such a hypothetical experiment, however, has not been realized yet because any imperfection in the optical lattice needed for suspending the atoms in Earth's gravity field would overwhelm the interference signal.

By comparison, standard atom interferometers such as the one used by Overstreet *et al.* are much less sensitive to imperfections of the laser fields, because each laser pulse is applied only for a short time and atoms are otherwise freely falling. The momentum kicks applied by vertical laser beams separate the two arms along the vertical direction. Therefore, the different gravitational time dilation experienced by atoms at different heights should lead to a proper-time difference between the upper and lower arm. Nevertheless, in a uniform gravitational field, this difference is exactly cancelled out by changes in kinetic energy caused by the gravitational acceleration, as can be understood by considering a freely falling reference frame (*10*). The interferometer outcome is thus entirely a consequence of the phase changes associated with the laser pulses, whose net contribution is proportional to the relative acceleration between the atoms and the light gratings.

By contrast, for sufficiently large arm separations, the effects of space-time curvature, which is linked to gravity gradients, lead to non-negligible deviations from a

Institute of Quantum Technologies, German Aerospace Center (DLR), Ulm, Germany. Email: albert.roura@dlr.de
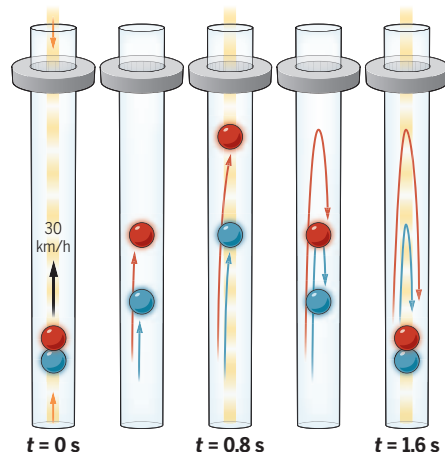
uniform gravitational field. Thanks to the large arm separation in the experiments performed by Overstreet *et al.*, the gravitational field of the massive ring at the top of the atomic fountain influenced the upper arm much more than the lower one, thus producing a measurable proper-time difference between the two. In principle, this difference could also be obtained by comparing two clocks following the same trajectories as the interferometer arms, but the difference would be far too small to be resolvable. Notably, interferometry experiments with atoms acting as quantum clocks can be sensitive to the substantially larger time dilation due to Earth's gravitational field by initializing the clock once the arms are spatially separated (*10*).

Measuring the effect of gravitational time dilation on matter-wave interference is a major step in the emerging field of gravitational quantum mechanics. Furthermore, the impressive sensitivity achieved in these experiments could be exploited in future measurements of Newton's gravitational constant (*3*), gravimetry applications (*11*), and tests of the universality of free fall (*12, 13*). Yet, important challenges remain because gravity gradients also lead to the dependence of the interference signal on the initial position and velocity of the atomic wave packets. This unwanted sensitivity to initial conditions is a major source of systematic uncertainties for precision measurements in nonuniform gravitational fields. Fortunately, a very effective technique to overcome these difficulties has recently been proposed (*14*) and is already playing a key role in high-precision tests of the universality of free fall (*13*). The prospects for improved measurements of Newton's gravitational constant based on atom interferometry are therefore very promising (*15*). ∎

REFERENCES AND NOTES
1. C. W. Chou, D. B. Hume, T. Rosenband, D. J. Wineland, *Science* **329**, 1630 (2010).
2. C. Overstreet *et al.*, *Science* **375**, 226 (2022).
3. G. Rosi, F. Sorrentino, L. Cacciapuoti, M. Prevedelli, G. M. Tino, *Nature* **510**, 518 (2014).
4. Y. Aharonov, D. Bohm, *Phys. Rev.* **115**, 485 (1959).
5. W. T. Lee, O. Motrunich, B. E. Allman, S. A. Werner, *Phys. Rev. Lett.* **80**, 3165 (1998).
6. H. Batelaan, A. Tonomura, *Phys. Today* **62**, 38 (2009).
7. T. Kovachy *et al.*, *Nature* **528**, 530 (2015).
8. K. Bongs *et al.*, *Nat. Rev. Phys.* **1**, 731 (2019).
9. M. A. Hohensee, B. Estey, P. Hamilton, A. Zeilinger, H. Müller, *Phys. Rev. Lett.* **108**, 230404 (2012).
10. A. Roura, *Phys. Rev. X* **10**, 021014 (2020).
11. A. Peters, K. Y. Chung, S. Chu, *Nature* **400**, 849 (1999).
12. D. Schlippert *et al.*, *Phys. Rev. Lett.* **112**, 203002 (2014).
13. P. Asenbaum, C. Overstreet, M. Kim, J. Curti, M. A. Kasevich, *Phys. Rev. Lett.* **125**, 191101 (2020).
14. A. Roura, *Phys. Rev. Lett.* **118**, 160401 (2017).
15. G. D'Amico *et al.*, *Phys. Rev. Lett.* **119**, 253201 (2017).

CELL BIOLOGY

Fetal bovine serum— a cell culture dilemma

Ethical and possible reproducibility issues arise when using fetal bovine serum in cell culture media

By **Jan van der Valk**

Fetal bovine serum [FBS, also known as fetal calf serum (FCS)] is a popular supplement to the basal medium used in cell and tissue culture. FBS is sourced from unborn calves at the slaughterhouse, raising ethical concerns about animal welfare. Recently, two different laboratories performed in vitro experiments that applied an identical experimental procedure and used cells and FBS from the same suppliers (*1*). The results they obtained were very different. Further analyses revealed that one cause for the difference in cell response was the supplementation of the cell culture medium with FBS, which had originated from different batches. Given the ubiquitous use of cell culture throughout research, it is important to ensure reproducibility as well as ethical sourcing of research products, such as the development of synthetic media.

To maintain and proliferate cells and tissues outside the body, an optimal environment with growth factors and nutrients is required. This is often a liquid medium. Since the first in vitro cell culture experiments that maintained frog nerve fibers in frog lymph fluid (*2*), there has been a search for the optimum medium composition. Over the years, both animal-derived media and artificial media were developed. Examples of artificial media are the Eagles medium and its improvements modified Eagles medium (MEM) and Dulbecco's MEM (DMEM), as well as Ham's F10, F11, and F12. Not all cells flourish in these artificial media. In 1958, it was discovered that cells can be maintained in active growth for longer periods of time in media containing FBS (*3*). Because it was also reported that several different human biopsies and primary cell cultures could successfully be proliferated in such supplemented media, this led to the widespread use of FBS that continues today.

FBS is naturally optimized for prenatal development of unborn calves, and it contains a wide range of nutrients and growth and adhesion factors in excess and has a low antibody content. Combined with its relatively inexpensive availability, FBS is historically the first choice for supplementing almost all eukaryotic cell culture media. Most cell types appear to respond well to FBS with regard to proliferation and viability. Unfortunately, two important observations from the first report of its use in cell culture have been largely overlooked: FBS can contain "toxic factors" that affect the quality of experiments, and the serum obtained in different seasons showed "appreciable variations of performance" [(*3*), p. 946].

That different batches of FBS, which are produced in different regions and during different times of the year, have different effects on cells is not surprising, given that it is a biological product with a largely unknown composition. This variance may become apparent in cell culture through effects on cell morphology, growth rate, and viability, as well as by altering responses in experimental settings (*1*). To avoid intralaboratory disparities in cell performance when a new batch of serum must be obtained, laborious batch-testing to check in-house quality criteria with the available cell lines in the lab are required. Although this solves in-house inconsistency, it does not address interlaboratory consistency, because different laboratories will not have access to the same batch of FBS. Interlaboratory reproducibility becomes crucial when in vitro methods are used in applied research, such as preclinical studies with human cells, or for regulatory safety testing of pharmaceuticals and chemicals. Because of these interlaboratory reproducibility issues, the Organisation for Economic Co-operation and Development (OECD) began to discourage the use of FBS in 2017, especially for human health risk assessments of chemicals (*4*).

Since 1989, the bovine spongiform encephalitis (BSE) crisis fueled efforts to replace bovine-derived material, particularly in clinical or pharmaceutical products. Because of the potential contamination with nonhuman pathogens, the risk of eliciting an unwanted immune response, and issues with product reproducibility, the US Food and Drug Administration (FDA) (*5*) and the European Medicines Agency (EMA) (*6*) discourage the use of FBS in cell and tissue culture for human clinical application.

There is a need to replace FBS supple-

mentation of tissue culture media with media that is more physiologically relevant, is safe to use, and has consistent composition. Different options, such as human or plant extracts or fully chemically defined media, are now available. FBS-free media are preferred for experiments involving human tissues and cells because FBS may contain zoonotic pathogens and xenogeneic proteins and so may be particularly incompatible (7). Where in vitro methods are used for clinical applications, xeno-free media, such as human serum or human platelet lysates, have been introduced. However, these are also biological products, and so there are batch-to-batch differences (8). In areas where reproducibility and safety are especially important, there is increasing attention on developing and using chemically defined media. Unlike FBS, chemically defined media are generally cell-type specific. Consequently, a different medium is used, or developed when not yet available, for specific cell types.

Several existing defined media from commercial sources or described in the scientific literature for a variety of cell types can be identified in the FCS-free database (https://fcs-free.org). For example, for Chinese hamster ovary (CHO) cells, which are often used in the production of recombinant therapeutic products, the development of suitable FBS-free media is extensively described (9). At present, 22 different FBS-free media for CHO cells are identified. At least 19 FBS-free media are available for human MRC-5 and primate Vero cells, which are widely used for vaccine production. Mesenchymal cells are isolated from human adipose cells for use in human regenerative therapies, and many FBS-free media have been developed for these purposes.

Synthetic media from commercial sources may have some drawbacks. Primarily, it is often unclear, for commercial and proprietary reasons, what the medium formulation is. It is therefore impossible to identify the relative amounts of specific components and thus their potential effects on cells and the substances to be tested. Furthermore, the composition may be changed, and commercial media are also expensive. In the ideal situation, the composition of the growth medium is published. This might give other researchers the opportunity to adapt media specifically for their cells. Although adaptations of existing FBS-free media for a particular cell type might cause interlaboratory irreproducibility, publication of the full formulations,
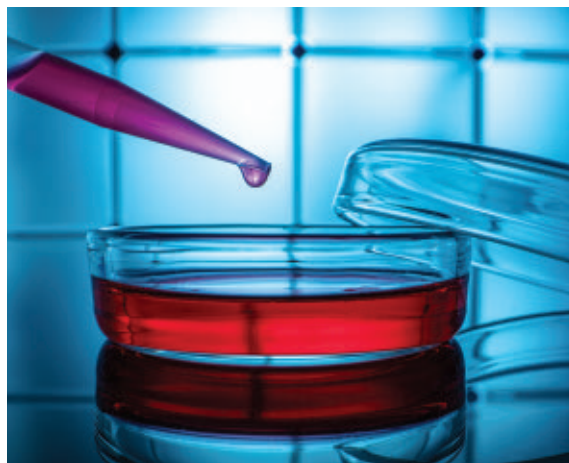
Cell culture medium is often supplemented with fetal bovine serum (FBS), which may be a source of experimental irreproducibility.

and comparing these, should allow the cause of any differences to be identified. This is impossible when using FBS and very difficult when using commercial medium. It is hoped that at some point, agreement is reached on the best medium for a particular cell type used for a specific purpose.

FBS-free media are not available for all cell types, so these have to be developed, or, ideally, a universal medium will be developed. The recommended starting point is a rich medium composed of equal volumes of DMEM and Ham's nutrient mixture F12, supplemented with recombinant insulin and transferrin and the mineral selenium (ITS). Other components, such as growth factors, hormones, and proteins, could further improve the medium (10). Furthermore, careful adaptation of the cells to the new medium (11), and the choice of substrate serving as the extracellular matrix, have been shown to be important to maintain or improve the phenotype of cells in an FBS-free environment (12). The cell type–specificity of chemically defined media poses a challenge for multi–organ-on-a-chip researchers when using different cell types within one system. Combining one chemically defined medium with a specific substrate for each cell type allowed a functional human multiorgan in vitro system, which could be maintained for 28 days (13).

There is not yet a universal chemically defined cell culture medium. The diversity of existing chemically defined media for some cell types and the work involved in developing those for other cell types are obstacles to replacing routine use of FBS. When new in vitro research models are developed without FBS, and agreements are reached between users about the best chemically-defined medium for a particular cell type and purpose, this will facilitate the acceptance and implementation of FBS-free cell culture.

In what medium cells and tissues express a phenotype that is physiologically relevant and appropriate for a particular study and how new media should be evaluated are questions that remain. Often, results obtained from cells grown in FBS-containing media are used as a reference for comparison. However, these results might not always be relevant and could lead to misinterpretation (14, 15). For human cell and tissue systems, comparison with results from cells cultured in human serum might be the solution, although, as with FBS, possible batch differences should be taken into account.

In vitro research is important to provide animal-free models that are physiologically relevant. Every potential cause for irreproducibility should be identified and a solution found. The use of FBS in cell culture medium is one of those causes, and the development of FBS-free media is crucial to ensure consistent, reproducible, and translatable results and safe products. ∎

REFERENCES AND NOTES

1. H. Barosova et al., Toxicol. In Vitro 75, 105178 (2021).
2. R. G. Harrison, M. J. Greenman, F. P. Mall, C. M. Jackson, Anat. Rec. 1, 116 (1907).
3. T. T. Puck, S. J. Cieciura, A. Robinson, J. Exp. Med. 108, 945 (1958).
4. OECD, "Guidance document on Good In Vitro Method Practices (GIVIMP) for the development and implementation of in vitro methods for regulatory use in human safety assessment" (OECD, 2017); www.oecd.org/env/ehs/testing/OECD_Draft_GIVIMP_in_Human_Safety_Assessment.pdf.
5. US Department of Health and Human Services (HHS), Food and Drug Administration, Center for Biologics Evaluation and Research, "Guidance for industry: Characterization and qualification of cell substrates and other biological materials used in the production of viral vaccines for infectious disease indications" (HHS, 2010); www.fda.gov/media/78428/download.
6. EMA, "Guideline on the use of bovine serum in the manufacture of human biological medicinal products" (EMA, 2013); www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/06/WC500143930.pdf.
7. N. Bryan, K. D. Andrews, M. J. Loughran, N. P. Rhodes, J. A. Hunt, Biosci. Rep. 31, 199 (2011).
8. K. Bieback, B. Fernandez-Muñoz, S. Pati, R. Schäfer, Cytotherapy 21, 911 (2019).
9. K. Landauer, in Animal Cell Biotechnology (Springer, 2014), pp. 89–103.
10. J. van der Valk et al., Toxicol. In Vitro 24, 1053 (2010).
11. M. S. Sinacore, D. Drapeau, S. R. Adamson, Mol. Biotechnol. 15, 249 (2000).
12. M. Cimino, R. M. Gonçalves, C. C. Barrias, M. C. L. Martins, Stem Cells Int. 2017, 6597815 (2017).
13. C. Oleaga et al., Adv. Funct. Mater. 29, 1805792 (2018).
14. A. J. Lau, T. K. Chang, Toxicol. Appl. Pharmacol. 277, 221 (2014).
15. Z. Wei, A. O. Batagov, D. R. F. Carter, A. M. Krichevsky, Sci. Rep. 6, 31175 (2016).

3Rs-Centre, Utrecht University, Utrecht, Netherlands. Email: j.vandervalk@uu.nl

PHOTO: LUCHSCHEN / ISTOCK.COM

The scent of atherosclerosis

Vascular macrophages sense an odorant to induce atherosclerotic plaque formation

By **Katey J. Rayner**[1,2] and **Adil Rasheed**[1,2]

I n the blood vessel wall, atherosclerotic plaque development (atherogenesis) occurs over several decades and involves the coordinated dysfunction of multiple cell types and organ systems. Although treatments targeting hypercholesterolemia and hypertension, the major risk factors of atherosclerosis, are effective at preventing fatal complications such as heart attack and stroke, there remains a major burden of atherosclerotic disease. Therefore, a more comprehensive understanding of the cells that drive disease, and how to therapeutically target them, is urgently needed. On page 214 of this issue, Orecchioni *et al.* (*1*) reveal an important role for olfactory receptor 2 (OLFR2) in atherosclerotic plaque development, identifying an additional target for intervention during the progression of atherosclerosis.

Olfactory receptors are a structurally diverse family of proteins that sense a wide range of odorants. The human genome comprises more than 400 olfactory receptors, which are expressed on the surface of olfactory sensory neurons in the nasal epithelium (*2, 3*). However, olfactory receptors are also expressed in sperm, skin, and the intestine, suggesting that they play a role outside of the olfactory system. Previous work identified the expression of *Olfr2* and its downstream signaling components in mouse vascular macrophages during atherosclerosis (*4*). Orecchioni *et al.* extend these observations and demonstrate that octanal, the endogenous ligand of the olfactory receptors, triggers calcium-dependent induction of the NLRP3 (NACHT, LRR- and PYD domains-containing protein 3) inflammasome and the production of proinflammatory interleukin-1β (IL-1β) to exacerbate atherosclerosis (see the figure).

Macrophages are present and maintain homeostasis in virtually every tissue, including the arterial wall, by promoting host defense and clearing deposited lipids and apoptotic cells. However, when these functions become disabled or overwhelmed, macrophages directly promote atherogenesis. The secretion of IL-1β by macrophages in the atherosclerotic plaque is a key process during atherosclerosis (*5, 6*), and there has been increasing clinical interest in blocking IL-1β in patients using monoclonal anti-

Atherogenesis involving octanal

Macrophages in the arterial wall are a hub of atherogenesis. Toll-like receptor 4 (TLR4) activation primes the NLRP3 (NACHT, LRR- and PYD domains-containing protein 3) inflammasome, which is fully activated by olfactory receptor 2 (OLFR2). Octanal, the OLFR2 ligand, is produced from oxidized low-density lipoprotein (oxLDL) cholesterol. This two-hit NLRP3 activation leads to interleukin-1β (IL-1β) production and exacerbation of atherosclerosis.
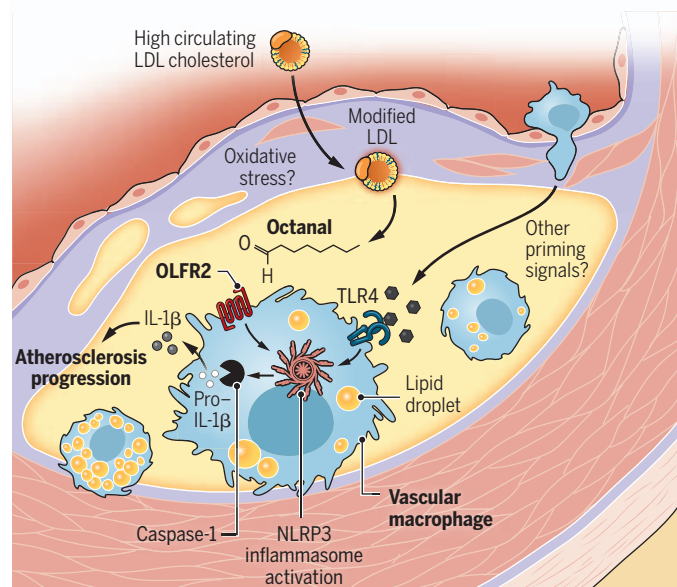
bodies (*7*). However, there are numerous issues associated with inactivating IL-1β, including increased risk of death from opportunistic infections. Given that octanal and OLFR2 act upstream of IL-1β, this could represent a therapeutic opportunity to block lipid-driven inflammation in the vascular wall without disarming host defense.

Octanal is a product of the peroxidation of fatty acids and occurs primarily as a result of oxidative stress in local vascular cells in the arterial wall (*8*). Orecchioni *et al.* posit that octanal is derived from oxidized low-density

lipoprotein (oxLDL), a modified form of LDL cholesterol, the so-called "bad" cholesterol that directly accumulates in the vessel wall and causes atherosclerosis (*5, 6*). The authors show that octanal signals through OLFR2 in macrophages, triggering downstream inflammation. Using aortic explants from mice with apolipoprotein E (*Apoe*) deletion (which increases production of cholesterol and atherogenic accumulation within macrophages), they showed that octanal could be produced by the aorta from oleic acid, a fatty acid prevalent in our diets. Notably, in mouse models of atherosclerosis and in humans, the circulating concentrations of octanal (~2 μM) are similar, suggesting that these mechanisms may translate to human disease. However, the precise source of octanal in the vessel wall is unknown.

Orecchioni *et al.* demonstrate that octanal is not sufficient to induce the activation of IL-1β by macrophages and that this induction of IL-1β only occurs with concomitant Toll-like receptor 4 (TLR4) agonism. TLR4 senses extracellular inflammatory molecules, leading to the transcriptional induction of NLRP3. This requirement for TLR4 ligation prior to octanal stimulation is reminiscent of previous studies demonstrating the two-hit activation of the inflammasome—priming and then activation—which is first driven by oxLDL engagement of TLR4 followed by signaling induced by cholesterol crystals that accumulate below the endothelial lining of the vessel wall during atherosclerosis (*9*). This raises the possibility that additional TLR priming, such as by other atherogenic ligands or by pathogen-associated molecular patterns (PAMPs), may act as a priming signal to activate the inflammasome in vascular macrophages. Thus, this study provides further evidence of the interplay of lipids and inflammation that occurs in macrophages and the role of macrophages as major immunometabolic hubs in atherosclerosis and other cardiometabolic diseases, including diabetes and obesity (*10*).

[1]University of Ottawa Heart Institute, Ottawa, ON, Canada. [2]Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada. Email: krayner@ottawaheart.ca

GRAPHIC: KELLIE HOLOSKI/SCIENCE

OLFR2 is expressed by multiple cell types within the aorta but was found to be most highly expressed in vascular macrophages (*1*, *4*). Although only ~30% of vascular macrophages expressed OLFR2 in mice, deletion of *Olfr2* or its downstream signaling mediators *Adcy3* (adenylate cyclase 3) and *Rtp1* and *Rtp2* (receptor transporter proteins 1 and 2) in hematopoietic cells led to a reduction in markers of plaque vulnerability (e.g., necrotic core in the vessel wall) that predispose to fatal complications, and an increase in collagen content (linked to plaque stability), indicating the potent role of OLFR2 signaling in macrophages on the atherosclerotic plaque. Moreover, despite the relatively low frequency of OLFR2⁺ vascular macrophages, delivery of exogenous octanal to atherosclerosis-prone *Apoe*-deficient mice led to about a twofold increase in plaque area. Together, these data indicate that octanal may signal through other olfactory receptors expressed by lesional cells and/or that OLFR2-expressing macrophages release proatherogenic signals that cross-talk with other cells in the plaque. This study, along with high-resolution single-cell sequencing of atherosclerotic plaques (*11*), further contributes to our understanding of the function of macrophage subpopulations and their influence on atherogenesis.

There remain some key unknowns regarding how octanal and its receptors function in the vessel wall. Is there another role for octanal, in addition to promoting inflammasome activation? It is possible that, similar to how odorants act as attraction or repulsion signals in the olfactory epithelium or sperm cells, octanal acts as a chemoattractant in the vessel wall. Are there additional sources of octanal, perhaps derived from diet or the environment? As large-scale omics studies continue to be used to identify additional genes associated with risk of cardiovascular-related death, and as their functions in the vessel wall are characterized, there is an opportunity to better understand atherosclerosis progression and derive new treatments. ∎

REFERENCES AND NOTES

1. M. Orecchioni *et al.*, *Science* **375**, 214 (2022).
2. C. Trimmer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9475 (2019).
3. B. Malnic *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2584 (2004).
4. S. McArdle *et al.*, *Circ. Res.* **125**, 1038 (2019).
5. K. J. Moore *et al.*, *Nat. Rev. Immunol.* **13**, 709 (2013).
6. P. Libby, *Nature* **592**, 524 (2021).
7. P. M. Ridker *et al.*, *N. Engl. J. Med.* **377**, 1119 (2017).
8. S. Janfaza *et al.*, *Biol. Methods Protoc.* **4**, bpz014 (2019).
9. A. Grebe *et al.*, *Circ. Res.* **122**, 1722 (2018).
10. A. Rasheed, K. J. Rayner, *Endocr. Rev.* **42**, 407 (2021).
11. A. Zernecke *et al.*, *Circ. Res.* **127**, 402 (2020).

Molten iron in Earth-like exoplanet cores

Iron crystallization in super-Earth interiors plays a key role in their habitability

By **Youjun Zhang**[1,2] *and* **Jung-Fu Lin**[3]

E arth, the only known habitable planet in the Universe, has a magnetic field that shields organic life-forms from harmful radiation coming from the Sun and beyond. This magnetic field is generated by the churning of molten iron in its outer core. The habitability of exoplanets orbiting other stars could be gleaned through better understanding of their iron cores and magnetic fields (*1*). However, extreme pressure and temperature conditions inside exoplanets that are much heavier than Earth may mean that their cores behave differently. On page 202 of this issue, Kraus *et al.* (*2*) used a powerful laser to generate conditions similar to those inside the cores of such "super-Earths" and reveal that even under extreme conditions, molten iron can crystallize similarly to that found at the base of Earth's outer core.

To date, more than 4500 exoplanets have been discovered, with approximately one-third of them categorized as Earth-like exoplanets (*3*). The discoveries of these exoplanets have raised hopes about finding habitable conditions beyond the Solar System and that exoplanetary habitability could be quite diverse in the Universe. Although surface water in a star's habitable zone has always been used as a qualifying condition for habitability, other key factors for habitability lie beneath the surface of the exoplanet, such as the property of its dynamo, a self-sustaining mechanism that generates a magnetic field (*4*).

Similar to Earth, super-Earths are thought to have formed through collisions and then differentiated into light silicate mantles and heavy iron cores. The iron cores were initially hot and molten but slowly lost heat to the silicate mantles. If core cooling is efficient, it can lead to iron crystallization, which releases energy. The cooling and solidification processes are thought to be the main sources of power that drives the convection of molten iron in the liquid core, generating magnetic fields through dynamo action, also known as magnetospheres. The pressure-temperature condition in which convection occurs is close to adiabatic, meaning that hot upwelling fluid follows a predictable temperature profile without heat gain or loss to the surroundings. Depending on the intersection relation between the iron melting temperature and the adiabatic profile under compression in a super-Earth's core, the molten cores can crystallize in two possible scenarios: either in an Earth-like "bottom-up" iron crystallization scenario or in an iron snowflake–like "top-down" scenario (see the figure). Bottom-up crystallization happens in the case of an iron melting curve steeper than the adiabatic pro-

Iron crystallization in super-Earth cores

Exoplanets with an Earth-like iron crystallization in their cores are more likely to possess and sustain a magnetic shield necessary for organic life-forms to exist. However, exoplanets with a higher content of light elements in the core may not have the internal condition necessary to sustain a solid core in the center and subsequently to sustain a magnetic shield over a long period.

Earth-like scenario

Silicate mantle

Solid iron

Molten iron

Convection

Magnetic field

Iron snowflake-like scenario

Iron crystals

Silicate mantle

Molten core
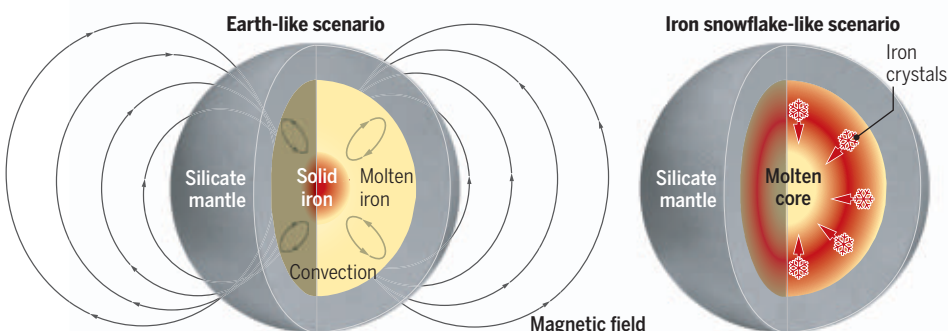
GRAPHIC: N. DESAI/*SCIENCE*

file, which is expected to be very efficient in powering and sustaining a dynamo, whereas core dynamos driven by an iron snowflake–like regime may be more difficult to maintain over a long period (5).

Experimental determination of the crystallization scenarios in super-Earths' cores is critical in assessing their magnetic fields and habitability. However, previous laboratory techniques have been limited to relatively low pressure-temperature ranges so that extrapolation to super-Earth cores and theoretical predictions were used in existing models (6). Kraus et al. used a laser to mimic the high pressure-temperature conditions and monitored iron crystallization up to ~1000 GPa and concluded that the Earth-like "bottom-up" scenario is the more likely outcome for super-Earth cores with iron-rich Earth-like compositions. This crystallization can promote the convection of molten iron to generate magnetic fields surrounding super-Earths more readily than previously thought.

Iron in Earth's core is under extreme pressures that range from 136 to 360 GPa and temperatures from 4000 to 6000 K. The melting curve of iron was previously determined up to ~300 GPa by using static and dynamic compression techniques (7–9). The advance of ultrahigh-power lasers (such as the National Ignition Facility) allows scientists to create much higher pressure and temperature conditions. Controlling the duration of the laser power allowed Kraus et al. to generate higher pressures and moderate temperatures to reproduce iron melting and crystallization processes at super-Earth core conditions.

The melting curve of iron up to ~1000 GPa determined by Kraus et al. indicates a melting slope steeper than the expected adiabat in a super-Earth's core. For a super-Earth with ~1.5 times the radius and ~5 times the mass of Earth, the melting temperature at its topmost outer core is estimated to be ~8500 K at ~600 GPa (2). Considering a silicate mantle temperature of ~5000 K at its bottom (10), a big temperature gradient across the super-Earth's core-mantle boundary could be expected. Therefore, a large heat flow and thermal energy source are responsible for powering its molten iron convection (11). As the super-Earth cools, its adiabat first intersects the melting curve of iron at its center, resulting in a bottom-up core solidification. This is the same crystallization scenario happening in Earth.

The thermochemical and gravitational energy provided by these processes can sustain convection and dynamo within super-Earths for billions of years (12). By contrast, the iron snowflake–like scenario can occur in the cores of planets and exoplanets with possible substantial amounts of light element(s) that would lower its melting curve. In the snowflake-like scenario, a cooling planet's adiabat intersects the iron melting curve near the top-middle of the core, leading to iron crystals forming and sinking toward its center. This scenario has been proposed to occur inside Mars because of its lower melting temperature caused by the presence of lighter element(s) in its core (5, 13).

When exoplanetary cores form, a certain amount of light elements—such as hydrogen, carbon, silicon, oxygen, and sulfur—make their way into the molten core (14). Their presence can depress the melting curve, influence the crystal structure stability of iron, and affect the output of thermochemical energy inside the core. Future experimental investigations of light element effects need to be taken into consideration in evaluating the dynamics of exoplanets at extreme conditions. Future investigation of the thermodynamic, transport, and rheological properties of silicate mantles and iron alloys at relevant super-Earth conditions can help us to better understand core dynamics, Earth-like mantle convection, and, potentially, plate tectonics. Detections of planetary magnetic fields outside of Earth's Solar System can be combined with laboratory measurements to infer exoplanetary interior processes and habitability. ∎

REFERENCES AND NOTES

1. T. Duffy, N. Madhusudhan, K. Lee, in *Treatise on Geophysics*, G. Schubert, Ed. (Elsevier, ed. 2, 2015), vol. 2, pp. 149–178.
2. R. G. Kraus et al., *Science* **375**, 202 (2022).
3. M. Mayor, C. Lovis, N. C. Santos, *Nature* **513**, 328 (2014).
4. B. J. Foley, P. E. Driscoll, *Geochem. Geophys. Geosyst.* **17**, 1885 (2016).
5. D. J. Hemingway, P. E. Driscoll, *J. Geophys. Res. Planets* **126**, e2020JE006663 (2021).
6. P. Driscoll, P. Olson, *Icarus* **213**, 12 (2011).
7. S. Anzellini, A. Dewaele, M. Mezouar, P. Loubeyre, G. Morard, *Science* **340**, 464 (2013).
8. S. J. Turneaure, S. M. Sharma, Y. M. Gupta, *Phys. Rev. Lett.* **125**, 215702 (2020).
9. J. Li et al., *Geophys. Res. Lett.* **47**, e2020GL087758 (2020).
10. C. T. Unterborn, W. R. Panero, *J. Geophys. Res. Planets* **124**, 1704 (2019).
11. Y. Zhang et al., *Phys. Rev. Lett.* **125**, 078501 (2020).
12. I. Bonati, M. Lasbleis, L. Noack, *J. Geophys. Res.: Planets* **126**, e2020JE006724 (2021).
13. S. C. Stähler et al., *Science* **373**, 443 (2021).
14. K. Hirose, B. Wood, L. Vočadlo, *Nat. Rev. Earth Environ.* **2**, 645 (2021).

[1]Institute of Atomic and Molecular Physics, Sichuan University, Chengdu, China. [2]International Center for Planetary Science, College of Earth Sciences, Chengdu University of Technology, Chengdu, China. [3]Department of Geological Sciences, Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, USA. Email: zhangyoujun@scu.edu.cn; afu@jsg.utexas.edu

NEURODEGENERATION

A molecular view of human amyloid-β folds

Structures of amyloid-β fibrils suggest Alzheimer's disease–modifying strategies

By **Michael Willem[1]** *and* **Marcus Fändrich[2]**

One of the mysteries of Alzheimer's disease (AD) etiology is the folding of the amyloid-β 42 (Aβ42) peptide, which forms aggregates ranging from small soluble and likely neurotoxic oligomers to mature amyloid fibrils that form amyloid plaques (1). Aβ peptides are derived from sequential cleavage of amyloid precursor protein (APP). Two main types of Aβ deposits can be distinguished in patient tissue: parenchymal amyloid plaques consisting mainly of Aβ42 and vascular amyloid deposits containing the shorter Aβ40 peptide (2). Previous research using cryo–electron microscopy (cryo-EM) determined the structures of Aβ40 fibrils from post mortem human AD brain tissue (3). On page 167 of this issue, Yang et al. (4) describe the cryo-EM structures of Aβ42 fibrils that were extracted from the brain tissue of patients with different neurodegenerative diseases, including AD. These structures aid in understanding the development of amyloid diseases and may inspire strategies for disease-modifying therapeutic intervention or diagnosis.

Yang et al. discerned three fibril morphologies—types I, Ib, and II. Type Ib represents a dimeric version of the type I fibrils. Type I fibrils were found primarily in sporadic AD patient material, whereas the type II filaments were mainly associated with familial AD patients and other neurodegenerative disorders (e.g., frontotemporal dementia), as well as being found in an amyloidogenic mouse model. The three fibril morphologies were assumed to have a left-hand twist, which corresponds to that of in vitro fibrils from Aβ42 or Aβ40 peptides but differs from the right-hand twist of Aβ40 fibrils from AD brain tissue

[1]Biomedical Center (BMC), Division of Metabolic Biochemistry, Faculty of Medicine, Ludwig Maximilians University Munich, Munich, Germany. [2]Institute of Protein Biochemistry, Ulm University, Ulm, Germany. Email: mwillem@med.uni-muenchen.de; marcus.faendrich@uni-ulm.de

(*3*). The peptide conformations in these newly described Aβ42 fibrils were S shaped at the carboxyl terminus and conformationally disordered at the amino terminus. Their conformations differ from the C-shaped peptide conformation and amino-terminal stability of Aβ40 fibrils from AD brain tissue (see the figure). Instead, the fibril peptides resemble the S-shaped structures and amino-terminal disorder of Aβ42 fibrils in vitro (*5–7*), although none of these structures exactly matched the peptide conformations reported by Yang *et al.*

This is in accordance with previous conclusions that in vitro fibrils do not necessarily model the specific morphologies of ex vivo fibrils (that were extracted from diseased tissue) (*3, 8–10*). Because ex vivo fibrils are typically more resistant to proteolysis than many in vitro fibrils, it has been suggested that disease-associated amyloid fibril structures arose by proteolytic selection in vivo (*8, 11*). However, a general concern of studying ex vivo fibrils is whether the extraction procedure might modify the fibril structure, specifically if this procedure involves 2% sarkosyl detergent, as in the case of Yang *et al.* Yet, type I and type II fibrils correlated with different neurodegenerative conditions, and details of their conformation differed from previously described in vitro structures (*5–7*), which indicates that the observed fibrillar morphologies are representative of the fibrils deposited in brain tissue.

What do these structures mean for the pathogenesis of AD? Most studies assume that prefibrillar oligomers of Aβ42, rather than Aβ42 fibrils, constitute the main pathogenic agents in AD (*1*). Notably, the S-shaped peptide fold in the fibrils described by Yang *et al.* resembles the S-shaped peptide conformation that was previously suggested for an Aβ42 oligomer (*12*). Genetic analyses of AD risk genes and longitudinal brain imaging studies plus ever-more-sensitive biomarker analyses indicate that the formation of fibrils and amyloid plaques occurs as early as 20 years before AD patients develop cognitive symptoms (*13*). After many failures of Aβ-directed therapeutic trials for AD, there is debate about the benefit of vaccination against aggregated Aβ. Therefore, gathering molecular insights on Aβ plaque composition is important to develop such vaccines. Hopefully, the molecular insights on Aβ structures (*3, 4*),

Amyloid fibril structures

Cryo–electron microscopy (cryo-EM)–derived structures of fibrils formed from amyloid-β40 (Aβ40) (*3*) and Aβ42 (*4*) from human brain tissue reveal different polymer conformations that are associated with distinct neurodegenerative diseases.

Aβ42
Type I
Gly⁹ — Ala⁴²

Type II
Val¹² — Ala⁴²

Aβ40
Asp¹ — Val⁴⁰
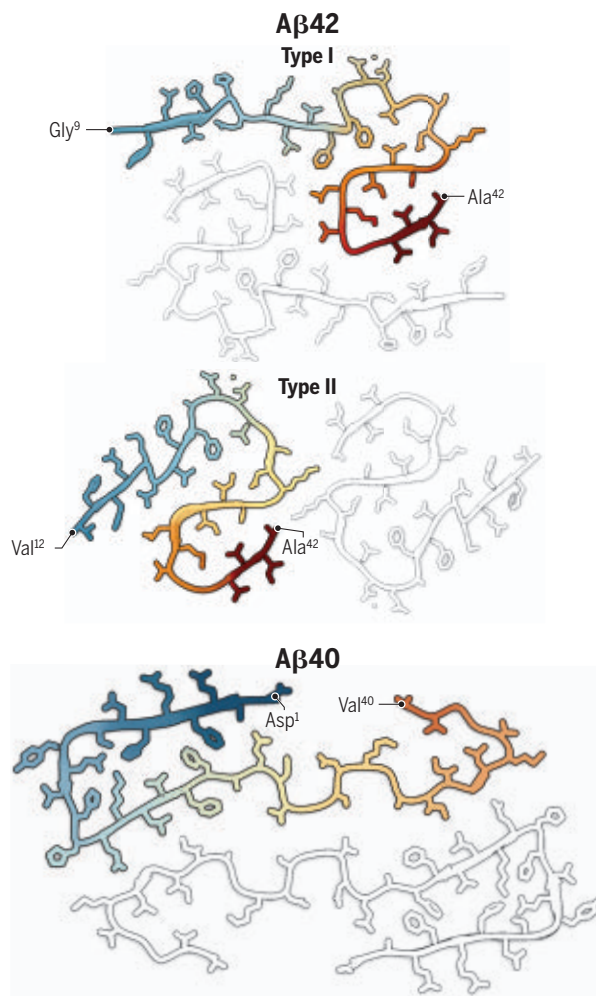
and similar knowledge obtained from other neurodegeneration-associated fibril structures, such as those derived from tau (*9*) and TAR DNA-binding protein 43 (TDP-43) (*10*), will aid in understanding the mechanism of neurodegenerative diseases and may generate new therapeutic or diagnostic strategies.

Antibodies are capable of reducing Aβ load in AD patients, but it remains unknown whether they are the right tools to break down oligomers and fibrils. The Aβ42 fibrils described by Yang *et al.* are conformationally disordered at the amino terminus, which harbors the epitopes targeted by many therapeutic antibodies, such as amino acids 3 to 7, recognized by the antibody aducanumab (*14*). The conformational disorder observed at this site indicates that this part of the fibrillar protein is particularly accessible for antibody binding. The conformational disorder may

stem, at least partly, from the numerous posttranslational modifications (PTMs) that affect the amino terminus of Aβ (*2*). PTMs are not included in the stable conformation of type I and type II Aβ42 fibrils, but detailed knowledge about the variability and the extent of specific PTMs in oligomeric Aβ and fibrillar Aβ will be instrumental for the development of inhibitors of fibril formation or drugs to break up amyloid plaques.

Molecular imaging with positron emission tomography with ligands that are based on the fibril-binding molecule thioflavin currently allows for the in vivo visualization and tracking of pathophysiological changes in AD and mild cognitive impairment (*15*). However, thioflavin derivatives may not differentially bind all relevant fibril types of Aβ, and therefore, new ligands with higher sensitivity and specificity are desirable for personalized medicine approaches. Moreover, Aβ seeding strategies, which allow the formation of amyloid fibrils in mice in a short amount of time, could be an alternative approach to test breakers of amyloid fibrils in vivo. But another important goal should be to determine the patterns of oligomers, protofilaments, and coaggregating molecules, such as apolipoprotein E, to better target them for disease prevention. The currently available structures of different ex vivo Aβ fibrils may open the door to develop such next-generation amyloid ligands. ∎

REFERENCES AND NOTES

1. C. Haass, D. J. Selkoe, *Nat. Rev. Mol. Cell Biol.* **8**, 101 (2007).
2. D. R. Thal *et al.*, *Acta Neuropathol.* **129**, 167 (2015).
3. M. Kollmer *et al.*, *Nat. Commun.* **10**, 4760 (2019).
4. Y. Yang *et al.*, *Science* **375**, 167 (2022).
5. Y. Xiao *et al.*, *Nat. Struct. Mol. Biol.* **22**, 499 (2015).
6. M. T. Colvin *et al.*, *J. Am. Chem. Soc.* **137**, 7509 (2015).
7. M. A. Wälti *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4976 (2016).
8. A. Bansal *et al.*, *Nat. Commun.* **12**, 1013 (2021).
9. S. H. W. Scheres *et al.*, *Curr. Opin. Struct. Biol.* **64**, 17 (2020).
10. D. Arseni *et al.*, *Nature* 10.1038/s41586-021-04199-3 (2021).
11. J. Schönfelder *et al.*, *Amyloid* **28**, 243 (2021).
12. M. Ahmed *et al.*, *Nat. Struct. Mol. Biol.* **17**, 561 (2010).
13. I. Sintini, J. L. Whitwell, *Curr. Opin. Neurol.* **34**, 525 (2021).
14. J. W. Arndt *et al.*, *Sci. Rep.* **8**, 6412 (2018).
15. P. Scheltens *et al.*, *Lancet* **388**, 505 (2016).

Inclusion in human–machine interactions

Human–machine interactions research should include diverse subjects and benefit all people

By **Tahira Reid** *and* **James Gibert**

Human–machine interactions (HMIs) describe how humans engage various systems, including those that are smart, autonomous, or both. Most HMIs either allow the human to control the machine (an instrument panel), allow the machine to obtain data (a heart monitor), or even both (a virtual reality setup). HMIs may be placed in three broad classes. In one class, the individual is active in the interaction—that is, the individual is the user or purchaser of a technology such as an automobile. In another class, the user is passive but consenting in the interaction—that is, the interaction occurs with their consent, such as the use of devices for medical diagnosis. There is also a class in which the user is passive and nonconsenting in the interaction, such as the use of facial recognition for law enforcement purposes.

The human in HMI implies inclusion regardless of ability, ethnicity, race, or social class, but the current research paradigm shows many examples of exclusion. The processes and procedures that govern HMI do not account for the heterogeneity that exists in human beings included in the data. However, a design science paradigm can inform the creation of the interfaces between humans and machines and their embedding in our natural, virtual, psychological, economic, and social environment. Efforts to overcome implicit biases being propagated through the HMI system require team members who can challenge cultural hegemony, along with great governance of design and operation.

These biases may manifest themselves in HMI in several ways when developing and implementing these systems. Imposing the cultural norms of an industrialized society in technology deployed in developing countries can lead to failure. For example, a more expensive recyclable product might still be disposed if the infrastructure in a developing country required for collection and transportation is not available and the users are not informed about the process (*1*). Differences between the cultures of

School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, IN, USA. Email: tahira@purdue.edu

nations can inhibit technology adoption, such as cultural assumptions in computer-user interfaces (*2*). Even within an industrialized country, the cultural norms of the upper socioeconomic class might accept a technology such as fingerprint, retinal, or voice recognition on a cell phone that could be deemed suspicious to lower socioeconomic class communities.

The cultural norms of an industrialized society often follow a traditional innovation process that targets high-income stakeholders versus an inclusive innovation process that also considers low-income stakeholders (*3*). These two approaches differ in how technologies are developed, whom they benefit most, and how stakeholders learn and engage with them. Urban air-mobility systems are a good example of how high-income communities are prioritized, such as in decisions about where airports are located in terms of noise and access, as evidenced by the income parameters included in their models (*4*).

In many countries, the focus is often on HMI where the individual is active in the interaction. Lower socioeconomic classes, which often include a greater fraction of ethnic minorities and immigrants, represent a demographic whose needs should be included in the design and trial phase of HMI systems. However, the population of individuals often reflected in HMI studies are commonly affluent and educated. The pressures of conducting research in a timely manner, either for publishing findings or delivering products, cause many researchers to work with study participants from campus or corporate communities out of convenience.

More severe consequences can result when the humans in an HMI system are subjects upon which data are taken—that is, they are passive and either consenting or nonconsenting in the interactions. In areas where subjects give consent, such as medical testing, biases in data recording and interpretation can result by not being intentional about human heterogeneity, both technically and culturally. For example, virtual doctor visits for skin conditions such as rashes or wound healing often are designed around white skin subjects and can fail for dark skin.

Biased algorithms do not work for all skin tones, ethnic groups, or genders (*5*).

Despite all of the advances in algorithm design, facial recognition algorithms have several known cultural problems in the basic identification of various faces from different races regardless of the country in which the algorithm is developed (see the figure). The dangers of bias and disregard of cultural and social considerations were identified by researchers at the National Institute of Standards and Technology (NIST) who noted that like humans, facial algorithms exhibit bias (*6*), also known as "algorithmic bias," according to digital activist Joy Buolamwini. As an example, algorithms developed by East Asian and Western countries identified faces from the races of developers more accurately than faces from other races.

Furthermore, recent data suggest that algorithms struggle to identify people of color. For example, ongoing work from NIST indicates that facial recognition software used by the Detroit Police Department misidentifies Black males between one and two times in 1000, and Black females between two and three times in 1000 (*7*). The numbers may appear insignificant but considering that an estimated 79% of Detroit's 4 million residents are Black, errors of this kind have led to at least one documented wrongful arrest (*7, 8*).

Deploying new technologies without engagement from key stakeholders is a prescription for failure. For example, from 2020 to 2021, the New York Police Department field tested a robotic dog (named "Digidog"), created by Boston Dynamics, in Brooklyn, Queens, and the Bronx. Although it was helpful in performing various tasks for this police force, the community in which it was embedded was not informed or educated about its use, functionality, and deployment (*9*). Employing this technology in communities is especially problematic given the historical use of dogs to capture slaves and as tools to terrorize civil rights protestors. The Digidog has been removed, but a greater threat lies on the horizon with a robot showcased at the annual Association of the United States Army, which has placed sniper rifles on robotic dogs developed by Ghost Robotics (*10*). Although the current program is confined to military use, it is conceivable that in the future it may be adapted for civilian police use. If it is adapted, the basis

of determining threat would likely not be clear to the public.

This question is not purely academic. In 2016, the Dallas Police Department ended a 5-hour standoff with a cornered sniper by using a robot that was modified to deliver a bomb that ultimately killed the sniper (*11*). This repurposing of the robot shows how, in certain situations, machines can perform dangerous duties and limit the loss of life of law enforcement, but this isolated case brings to the forefront the question of when machines should be used in this capacity. The rules of conduct and use that govern their public interactions must consider adverse effects on minority communities. Laws, regulations, and policies have to be created or updated to address their use and any unintended consequences.

In the case of this Digidog, the algorithms may be capturing variables such as race, gender, ethnicity, and other information that could create and reinforce biased data, leading to discriminatory practices (*12*) and violating situational fairness (*13*). In the case of facial recognition algorithms, the pool of training data may be heavily biased toward one race, which would lead the algorithm to make false predictions when it encounters other races.

Despite these ongoing limitations, culturally aware HMI is possible, of which we discuss three fundamental aspects. One is to identify ways to partner directly with communities of interest. For example, members of a low-income community became partners in the research with leadership roles (*14*). This model has transformative potential for HMI researchers such that communities not typically represented in HMI data can foster empathy and be part of the initial discussion, as opposed to being consulted after product development or excluded completely. Consider 10 years from now when an autonomous ride-sharing vehicle is sent in response to a ride request placed by a relative or a social worker to pick up passengers with little education and no prior exposure to a driverless vehicle. The riders would need to be prepared beforehand and ensured of their safety.

Another aspect of culturally aware HMI is to recruit diverse students to engage in this research who can also serve as ambassadors to diverse communities. Despite the great pressures to publish research quickly, constant reliance on convenience samples for datasets is problematic and exclusionary. Researchers could recruit and train students as early as high school on best practices in human-subject studies and partner with them to reach diverse communities. For example, one research team

Face-recognition disparities

Similarity is evaluated as a percentage between photos of a white male and a Black female. In the white male face, the similarity is higher. For the Black female face, the similarity is lower but still concluded (incorrectly) that a positive identification has been made.
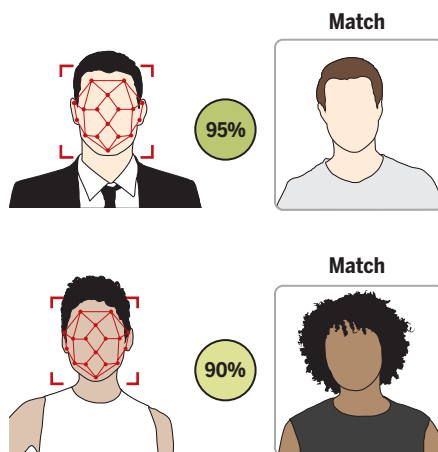
worked with minoritized high school students to solve problems related to aerial drones, which involved natural interactions with members of their own community (*15*).

Culturally aware HMI is also possible by instituting third-party oversight (for example, review boards). Its role is to provide unbiased accountability for how HMIs are used and deployed to be fair toward and ensure the safety of humanity, and that diverse participants are included in the studies. This approach would mirror requirements in pharmaceutical testing, which involves ethical participation, informed consent in clinical studies, and preregistration of diverse participants to ensure efficacy and safety across the general population. For the academic community, funding agencies would need to require such participation and build the costs of recruiting and remunerating participants into grants. For academic HMI researchers, it is important to acknowledge the lack of diversity to date in most human trials, explain the moral and financial benefit of diversity and inclusion, and let their own research efforts lead by example.

For small start-up companies, the success of a new product is not ensured, and testing is often limited by cost to specialized "beta groups." However, a diverse design team is more likely to build in features that make an HMI more accessible, and testing with diverse subjects earlier in the process can identify design flaws. However, greater inclusivity of participants is clearly needed for products targeting public applications aimed at surveillance and security, as well as medical applications.

The developers of technological ad-

vances for the general public must mitigate harm, not only by minimizing risks (such as safety features on vehicles) but also by educating the public so that they can make informed decisions, similar to patients deciding that the side effects of a drug outweigh its benefits. In the pharmaceutical industry, immoral overreaches to increase profits can have severe financial consequences because of regulatory oversight by government regulators. Similar regulatory efforts could be applied in HMI research to ensure a level playing field, but also carry the risk that only large established companies could develop products.

The need to ensure equity is clear, although the path to it is not. The drive to develop products for high-income users and the current lack of diversity argue that private-sector efforts alone are insufficient. Regulatory oversight is needed that considers the financial, safety, and privacy rights of participants. Such efforts require that the general public, academia, and industrial communities develop guidelines to ensure that HMIs that are safe, equitable, and accessible regardless of cultural or socioeconomic differences. ∎

REFERENCES AND NOTES

1. L. Salgado, R. Pereira, I. Gasparini, in *Human-Computer Interaction: Design and Evaluation*, M. Kurosu, Ed. (Springer International, 2015), pp. 60–70.
2. Z. Ishak, A. Jaafar, A. Ahmad, *Procedia Soc. Behav. Sci.* **65**, 793 (2012).
3. C. Foster, R. Heeks, *Eur. J. Dev. Res.* **25**, 333 (2013).
4. P. D. Vascik, R. J. Hansman, N. S. Dunn, *J. Air Transport.* **26**, 133 (2018).
5. A. Howard, Sex, Race, and Robots: How to Be Human in the Age of AI (Audible Originals, 2020).
6. P. Grother, M. Ngan, K. Hanaoka, Face Recognition Vendor Test (FRVT): Part 3, Demographic Effects (National Institute of Standards and Technology, 2019).
7. K. Wiggers, "NIST benchmarks show facial recognition technology still struggles to identify Black faces," *VentureBeat*, 9 September 2020.
8. B. Allyn, "'The Computer Got It Wrong': How Facial Recognition Led To False Arrest Of Black Man," *NPR*, 24 June 2020.
9. S. Bushwick, *Sci. Am.* (7 May 2021); www.scientificamerican.com/article/the-nypds-robot-dog-was-a-really-bad-idea-heres-what-went-wrong/.
10. A. Selyukh, "Bomb Robots: What Makes Killing In Dallas Different and What Happens Next?" *NPR*, 8 July 2016.
11. D. Hambling, *New Scientist* (14 October 2021).
12. A. Howard, J. Borenstein, *Sci. Eng. Ethics* **24**, 1521 (2018).
13. L. Floridi, J. Cowls, T. C. King, M. Taddeo, *Sci. Eng. Ethics* **26**, 1771 (2020).
14. S. L. Syme, *Prev. Chronic Dis.* **1**, 1 (2004); www.cdc.gov/pcd/issues/2004/jan/03_0001.htm.
15. J. Bhuyan *et al.*, *TechTrends* **64**, 899 (2020).

LINGUISTICS

An antiquated language, reimagined

A new tome traces efforts to unify, reform, and modernize the Chinese language

By **Zuoyue Wang**

As China's scientific, technological, and economic developments continue to propel its rapid rise and geopolitical tensions, there is a hunger for information on the country's past and present. Among the new crop of books being published to meet this demand, *Kingdom of Characters* by Yale professor of East Asian languages and literatures Jing Tsu stands out as a lively and insightful history of the intersection of China's information technology systems and its language revolution. The book is a richly documented, riveting, and scholarly rigorous transnational account of how Chinese evolved from a hard-to-learn script entrenched in the beleaguered Middle Kingdom in the 19th century to a global language in the 21st century.

Tsu profiles a number of historical and contemporary figures who represent successive campaigns, begun around the turn of the 20th century, to unify, reform, and modernize the Chinese script. In doing so, she provides not only vivid portraits of these individuals and their struggles but also broad coverage of social and political changes in modern Chinese history.

The narrative begins with Wang Zhao, a persecuted political reformer who, disguised as a monk, snuck back into China from his exile in Japan in 1900 to pursue another radical agenda: the unification of numerous Chinese dialects into what became known as Mandarin and the invention of a new phonetic alphabet for Chinese characters. Both efforts were aimed at improving literacy among the Chinese people, a goal shared by other reformers.

The 1911 republican revolution that ended the Qing imperial rule of China not only lent support to Wang's Mandarinization drive but

also intensified nationalism, ultimately inspiring Zhou Houkun and Qi Xuan, two Chinese students studying in the US, to invent Chinese typewriters with rival designs in the mid-1910s. The enormous quantity of Chinese characters and the difficulty of constructing them satisfactorily from their component
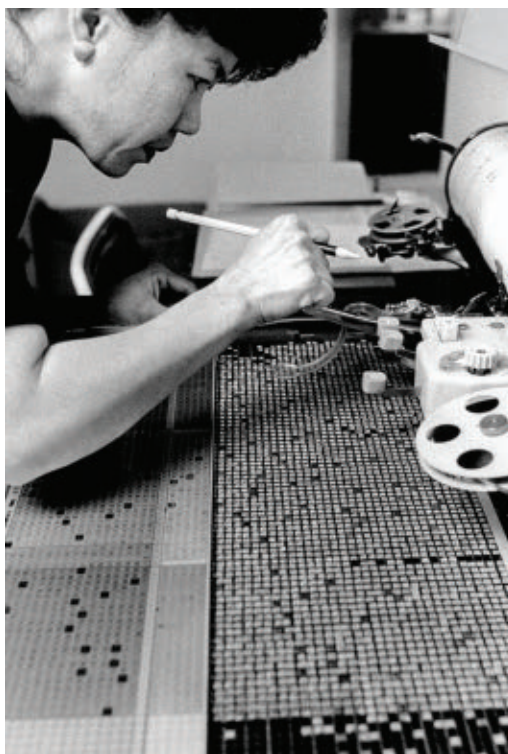
Lee Mau Yong operates a Chinese typewriter used to set type for printing in 1986.

parts proved challenging, but an improved model based on Zhou's design did come into wide use beginning in the 1920s (*1*).

The Chinese script's incompatibility with the Western alphabet not only posed cultural and technical barriers but also resulted in higher prices for Chinese users of telegraphy, who were forced to use the more expensive numerals to render Chinese characters. The battle to invent a suitable symbolic system to

replace or accompany Chinese script raged on during the 1930s and 1940s.

When the Chinese Communist Party drove the nationalist government to Taiwan and established the People's Republic in 1949, the reform entered a decisive new phase with two changes that have reshaped it to this day: simplified characters and the development of pinyin as the official alphabet designed to accompany the script. In an original discussion of the roots of pinyin, Tsu uncovers both its connections with Soviet linguistics and the overlooked roles played by a group of Chinese Muslims who were involved in language reforms in both China and Soviet Kyrgyzstan.

As the account moves closer to the present, Tsu describes how the Chinese language was brought into the digital world against the background of the chaotic Cultural Revolution, the modernization drive in the ensuing reform era, and the reopening of relations between China and the US. Here, she profiles Zhi Bingyi, an electrical engineer who was placed in solitary confinement in 1968 when he figured out a way to enter Chinese characters into a computer, and Wang Xuan, a computer researcher who led efforts to solve the critical problem of character compression in the mid-1970s.

A critical step in Wang's achievement, we learn, pivoted on design information about an advanced new microchip provided in 1979 by Francis F. Lee, a Chinese American engineer. Tsu reconstructs Lee's deep, but little-known, involvement in digitizing Chinese characters in association with an American nonprofit known as the Graphic Arts Research Foundation.

Tsu devotes the last chapter of *Kingdom of Characters* to the globalization of the Chinese language. In the book's conclusion, she recounts a contentious 2018 conference in Hanoi, Vietnam, where a Unicode group adjudicated new Chinese characters proposed for international acceptance. This episode illustrates a larger point: that even as the Chinese language goes global, it still stirs cultural and geopolitical tensions, especially in an era when China seeks to expand its influence in the world. ∎

The reviewer is at the Department of History, California State Polytechnic University, Pomona, CA 91768, USA. Email: zywang@cpp.edu

REFERENCES AND NOTES
1. For more on the history of the Chinese typewriter, see T. S. Mullaney, *The Chinese Typewriter: A History* (MIT Press, 2017).

10.1126/science.abn0932

PHOTO: *DENVER POST*/GETTY IMAGES

DEMOGRAPHY

Who gets counted and how

Sexual orientation and gender identity data must be collected and used with thought and with care

By **Nancy Bates**

The provocative word "queer" in the title of this book sets the stage for what is to come: controversy, nuances, and all the "messiness" that comes with attempts to measure subpopulations of people who identify as sexual or gender minorities. Organized into three parts—data collection, data analyses, and the use of queer data—the book does an admirable job explaining the finer points behind the complicated constructs of sexual orientation and gender identity (SOGI) and drawing attention to nuances that make it difficult to precisely measure micro-minorities.

measure sexual orientation, gender identity, and transgender history for the first time. Here, he delves into the complex web of decisions that determined the nature of the census questions used to measure lesbian, gay, bisexual, transgender, and queer (LGBTQ) populations, revealing, for example, how seemingly mundane phrasing, response options, and even instructions can greatly affect who gets counted and how they are classified.

Guyan argues that the census ultimately risks undercounting the country's LGBTQ populations, as many people are "unable to participate or forced to identify in a way that does not accurately reflect who they are." And even seemingly minor mistakes committed

the US Census Bureau sponsored a nationwide advertising campaign and hired LGBTQ partnership specialists to serve as "trusted voices" to educate and increase participation among sexual and gender minorities.

In chapter 4, "Beyond Borders: Queer Data Around the World," Guyan offers an overview of how language and cultural differences influence the measurement of LGBTQ populations globally. Some countries, he notes, are moving away from mandatory collection of sex data on legal documents and allowing instead for a nonbinary gender designation. Meanwhile, New Zealand's new "gender by default" principle defaults to the collection of gender data as opposed to sex. Here, Guyan misses an opportunity to encourage stakeholders to push for the collection of SOGI items in the European Social Survey, which does not currently collect such items.

In chapter 5, Guyan argues that "the cleaning of data can remove its queerness." Analysts may decide to combine several response categories to achieve sufficient case numbers for statistical testing, for example, a move that serves a practical purpose but can also mask differences that exist between subpopulations. Analysts may also remove answers that do not conform with conventional questionnaire design, such as when respondents strike out "male" and "female" and write "nonbinary" in the margins.

The book concludes with eight "queer questions" for practitioners to consider as they engage in the collection, analysis, and application of SOGI research: "Do your methods present an authentic account of LGBTQ lives?" for example, and "Do you elevate LGBTQ lives and critically examine the invisibility of majority characteristics?"

Queer Data lacks any substantive discussion around collecting data on intersex populations. Measures of this population are almost never collected, yet intersex is increasingly being included in the sexual minority "umbrella." Nonetheless, the book is a welcome addition on a topic that currently lacks wide attention. Guyan poses provocative questions that practitioners should consider before embarking on research that focuses on sexual and gender minorities. ∎

Activists in Edinburgh respond to individuals protesting changes to Scotland's census in September 2021.

Throughout the book, Guyan highlights how decisions made during different stages of the data life cycle are seldom value-neutral and can introduce biases to the final outcomes. This parallels the total survey error approach to understanding the life cycles of surveys (*1*). A nod to this literature could have provided a more holistic view and one more grounded in survey methods theory.

In chapter 3, perhaps the book's strongest, Guyan offers an insider's view of Scotland's 2022 census as the government prepares to

by sexual and gender majorities can lead to disproportionately large measurement errors. For example, adding response options beyond gay, lesbian, and bisexual (e.g., pansexual or queer) may cause confusion among cisgender and straight individuals, potentially resulting in significant overestimations of minority populations.

People who identify as LGBTQ may also be fearful of how the information will be used or shared. I was disappointed that the book failed to offer any discussion regarding outreach strategies that may or may not have been used to mitigate distrust and increase engagement and participation in Scotland's census. In advance of the 2020 US census,

The reviewer is a retired senior survey methodologist, US Census Bureau, Suitland, MD 20746, USA. Email: batesnancy5@gmail.com

REFERENCES AND NOTES

1. R. M. Groves *et al.*, *Survey Methodology* (Wiley, ed. 2, 2009).

10.1126/science.abn1911

Pakistan's flare-horned markhor (*Capra f. falconeri*) is once again at risk.

Edited by **Jennifer Sills**

Pakistan's markhor population in decline

The flare-horned markhor (*Capra f. falconeri*), a subspecies of the national animal of Pakistan, is native to the high mountain landscapes in northern Pakistan (*1*). In 2015, the International Union for Conservation of Nature (IUCN) downlisted the species from Endangered to Near Threatened (*2*), a reasonable decision at the time; between 2012 and 2016, the flare-horned markhor increased from 1900 to 2800 individuals (*1*). However, the population has drastically decreased since then. Urgent steps must be taken to accurately assess the markhor's status, update the IUCN listing, and reverse the recent population decline.

The flare-horned markhor population grew to 2868 individuals in 2019 but has since dropped to fewer than 800 (*3*). The lack of funding for local national park employees who monitor wildlife (known as watchers) has resulted in unchecked illegal hunting by those who use or sell markhor meat. Feral dogs have killed 392 markhors in the Chitral Gol National Park in the past 15 years (*4*). The negligence of the government and local communities and the lack of resources and scientific research have impeded markhor conservation efforts.

Urgent action is needed while the markhor population's survival remains possible. The Pakistani government should immediately reinstate the salaries of watchers (*5*), strictly monitor illegal hunting of the markhor, regulate the population of feral dogs, and raise awareness in the local community. Trophy hunting, which was legalized when the markhor population was higher (*6*), should now be banned unless an independent survey confirms stable population numbers. Finally, the wildlife department should allow national and international researchers to conduct independent surveys to better conserve this mountain monarch.

Shahid Ahmad and Ghulam Nabi*
Key Laboratory of Animal Physiology, Biochemistry, and Molecular Biology, College of Life Sciences, Hebei Normal University, Hebei, Shijiazhuang 050010, China
*Corresponding author.
Email: ghulamnabiqau@gmail.com

REFERENCES AND NOTES

1. L. Adhikari *et al.*, *Environ. Challenges* **11**, 100175 (2021).
2. S. Michel, T. Rosen Michel, "*Capra falconeri* (errata version published in 2016)" (IUCN Red List of Threatened Species, 2015).
3. Mukhamuddin, "Markhor population dwindling in Chitral," *The Express Tribune* (2021); https://tribune.com.pk/story/2326810/markhor-population-dwindling-inchitral.
4. R. H. Khattak *et al.*, *Braz. J. Biol.* **20**, 83 (2021).
5. H. Hayder, "Markhor population in Chitral drops 35% since 2019 due to KP govt's decisions," ProPakistani (2021); https://propakistani.pk/2021/10/29/markhor-population-in-chitral-drops-35-since-2019-due-to-kp-govts-decisions/.
6. W. Rashid, J. Shi, S. Dong, H. Sultan, *Animals* **10**, 597 (2020).

Central nervous system weapons dealt a blow

Since its adoption in 1993, the Chemical Weapons Convention has banned the development, possession, and use of weaponized toxic chemicals (*1*). However, whether this prohibition also applied to law enforcement use of certain agents that act on the central nervous system (CNS) remained the subject of debate. In December 2021, after years of applied scientific research [e.g., (*2–5*)], public awareness raising, and advocacy, the Conference of States Parties to the Chemical Weapons Convention adopted a landmark Decision to effectively outlaw the aerosolized use of CNS-acting chemical agents for law enforcement purposes (*6*). This is welcome news for those who want to ensure that benignly intended scientific research is not misused for hostile purposes, but the policy leaves room for improvement.

Although 85 countries supported the Decision, including Australia, Switzerland, and the United States, the vote was opposed by 10 countries, which may not feel constrained by its prohibitions (*7*). Notable among the opponents was Russia, whose security forces used aerosolized fentanyl derivatives to end the 2002 Moscow theater siege, causing the deaths of more than 120 hostages from poisoning and asphyxiation (*8*). Subsequent dual-use research into CNS-acting chemicals has been reported by Russian scientists as well as scientists from China and Iran, who also opposed this Decision (*9*).

Furthermore, the Decision is limited in scope. It explicitly prohibits only aerosolized CNS weapons, excluding other delivery mechanisms such as law enforcement dart guns (*9*). Similarly, future

law enforcement weapons that use toxic chemicals (including toxins and bioregulators) acting on other human physiological processes would not be covered. The International Committee of the Red Cross has argued that the Decision should not be interpreted as meaning that such weapons are acceptable in law enforcement (*10*).

This Decision is a success to be celebrated. However, the task now falls on the Chemical Weapons Convention States to implement it fully. States must clarify the range of toxic chemicals covered by this Decision; identify and verifiably destroy relevant CNS-agent weapon stocks, munitions, and other delivery mechanisms; and close down research and development of new CNS-acting weapons. It is vital that the medical and scientific communities continue their independent oversight to ensure that countries take these steps. Life and chemical sciences should never be weaponized again.

Michael Crowley and Malcolm Dando*
Section of Peace Studies and International Development, University of Bradford, Bradford, West Yorkshire BD7 1DP, UK.
*Corresponding author.
Email: mrdando@bradford.ac.uk

REFERENCES AND NOTES

1. "Convention on the Development, Production, Stockpiling and Use of Chemical Weapons" (Organisation for the Prohibition of Chemical Weapons, 1993).
2. British Medical Association, "The use of drugs as weapons" (2007); https://wellcomecollection.org/works/axh7rs2q.
3. Spiez Laboratory, "Technical Workshop on Incapacitating Chemical Agents" (Spiez, Switzerland, 2012).
4. "Brain Waves Module 3: Neuroscience, conflict and security" (Royal Society, 2012).
5. M. Crowley, M. Dando, L. Shang, Eds., *Preventing Chemical Weapons: Arms Control and Disarmament as the Sciences Converge* (Royal Society of Chemistry, London, 2018).
6. Organisation for the Prohibition of Chemical Weapons, "Decision: Understanding regarding the aersolised use of central nervous system-acting-chemicals for law enforcement purposes" (CWC Conference of States Parties, 26th Session, 2021).
7. CWC coalition, "26th Session of the Conference of the States Parties (CSP-26)" (2021); www.cwccoalition.org/csp26-summary/.
8. J. Riches, R. Read, R. Black, N. Cooper, C. Timperley, *J. Anal. Toxicol*. **36**, 647 (2012).
9. M. Crowley, M. Dando, "Down the slippery slope? A study of contemporary dual-use chemical and life science research potentially applicable to incapacitating chemical agent weapons" (Bath University, 2014).
10. International Committee of the Red Cross, "Statement by the ICRC, 26th Session of the Conference of the States Parties (CSP-26)" (2021); www.icrc.org/en/document/prohibition-chemical-weapons-26th-conference.

Transboundary conservation's rise

In their Letter "Poland's border wall threatens ancient forest" (26 November 2021, p. 1063), B. Jaroszewicz *et al*. discussed negative ecological impacts of a border wall that Poland has started to build for political reasons. As they explain, border walls are bad for biodiversity (*1*), and yet wall constructions are increasing globally (*2*). Although we agree that borderlands are in trouble, there is a positive sign: Global transboundary conservation is also on the rise.

Transboundary conservation is a top priority for biodiversity conservation (*3*). Theoretical and empirical studies have recently explored the effects of political borders on genetic, behavioral, species, and ecosystem changes (*1–4*) and proposed conservation planning for border regions under climate and political changes (*5*, *6*). In addition, systematic biodiversity inventories have been compiled in borderlands that were previously inaccessible to researchers. Thousands of new species, often threatened, have been discovered, helping to promote practical transboundary conservation actions (*3*).

For example, to protect the Critically Endangered Cao-vit gibbon (*Nomascus nasutus*), governments of both Vietnam and China established protected areas to facilitate effective transboundary conservation. As a result, the population size of this species increased from between 102 and 110 in 2007 to between 122 and 134 in 2018 (*7*). Similarly, the establishment of a 15,000 km² national park in the China–Russia border region promoted the recovery of Siberian tigers and Amur leopards (*8*). Globally, the number of transboundary protected areas has substantially increased (*9*).

Public concerns about the construction of walls, wildlife trafficking, armed conflicts, and their associated impacts on wildlife in borderlands could be influential in political decisions, to the benefit of these ecosystems (*10*). Such public support has already motivated politicians in some regions to implement influential transboundary conservation programs. For example, the Heart of Borneo initiative, signed in 2007 by the governments of Brunei, Indonesia, and Malaysia, has facilitated transboundary cooperation to protect 220,000 km² of tropical rainforest (*11*). The Leticia Pact, signed in 2019 by seven South American Amazonian countries, aims to collaboratively protect the world's largest rainforest (*12*). If more transboundary conservation efforts are made, the negative effects of international borders on biodiversity could be avoided or substantially reduced. We hope that this conservation impetus will overcome recent trends to fortify borders with linear, uncrossable structures.

Jiajia Liu[1]*, Tiantian Zhang[1], Luke Gibson[2]
[1]Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Biodiversity Science, School of Life Sciences, Fudan University, Shanghai 200438, China. [2]School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China.
*Corresponding author.
Email: liujiajia@fudan.edu.cn

REFERENCES AND NOTES

1. N. Fowler, T. Keitt, O. Schmidt, M. Terry, K. Trout, *Front. Ecol. Environ*. **16**, 137 (2018).
2. J. D. C. Linnell *et al.*, *PLOS Biol*. **14**, e1002483 (2016).
3. J. Liu, D. L. Yong, C. Y. Choi, L. Gibson, *Trends Ecol. Evol*. **35**, 679 (2020).
4. A. Trouwborst, F. Fleurke, J. Dubrulle, *Rev. Eur. Comp. Int. Environ. Law*. **25**, 291 (2016).
5. L. Wang *et al.*, *Conserv. Biol*. **35**, 1797 (2021).
6. M. A. Titley, S. H. M. Butchart, V. R. Jones, M. J. Whittingham, S. G. Willis, *Proc. Natl. Acad. Sci. U.S.A*. **118**, e2011204118 (2021).
7. C. Ma *et al.*, *Oryx*, **54**, 776 (2019).
8. K. McLaughlin, *Science* 10.5555/article.2404560 (2016).
9. I. Lysenko, C. Besançon, C. Savy, "2007 UNEP-WCMC global list of transboundary protected areas" (2007); www.sadc.int/files/6313/5852/1201/unep-2007-global-list-of-transboundary-pas-en.pdf.pdf.
10. R. Peters *et al.*, *Bioscience* **68**, 740 (2018).
11. M. Lim, *Asia Pacific J. Environ. Law* **17**, 65 (2014).
12. P. R. Prist *et al.*, *Science* **366**, 699 (2019).

TECHNICAL COMMENT ABSTRACTS

Comment on "Evidence of humans in North America during the Last Glacial Maximum"
David B. Madsen, Loren G. Davis, David Rhode, Charles G. Oviatt
Bennett *et al*. (Reports, 24 September 2021, p. 1528) report human footprints from Lake Otero, New Mexico, USA, ~22,000 years ago. Critical assessment suggests that their radiocarbon chronology may be inaccurate. Reservoir effects may have caused radiocarbon ages to appear thousands of years too old. Independent verification of the ages of the footprint horizons is imperative and is possible through other means.
Full text: dx.doi.org/10.1126/science.abm4678

Response to Comment on "Evidence of humans in North America during the Last Glacial Maximum"
Jeffrey S. Pigati, Kathleen B. Springer, Matthew R. Bennett, David Bustos, Thomas M. Urban, Vance T. Holliday, Sally C. Reynolds, Daniel Odess
Madsen *et al*. question the reliability of calibrated radiocarbon ages associated with human footprints discovered recently in White Sands National Park, New Mexico, USA. On the basis of the geologic, hydrologic, stratigraphic, and chronologic evidence, we maintain that the ages are robust and conclude that the footprints date to between ~23,000 and 21,000 years ago.
Full text: dx.doi.org/10.1126/science.abm6987

RESEARCH

EVOLUTIONARY BIOLOGY
Coral adaptation in the world's warmest seas

Corals within the Persian and Arabian Gulf, a shallow inland sea flooded during the Holocene, can withstand the highest temperatures known for coral reefs today (37°C) and have a distinctive population structure across the basin. These corals evolved extraordinary thermal tolerance within 10,000 years. Smith *et al.* examined portions of the genome and epigenome of the brain coral *Platygyra daedalea* across the Gulf. This work reveals thermal tolerance genotypes that will be critical to preserve in the quest to save coral reefs from climate change. —MM   *Sci. Adv.* 10.1126/sciadv.abl7287 (2022).

A genomic survey of brain corals, pictured here, reveals genotypes associated with recently evolved thermal tolerance in some populations.

TOPOLOGICAL MATTER
Teasing out the topological order

Quantum Hall states, which form in two-dimensional electron gases at low temperatures and in the presence of strong magnetic fields, have long been known to have nontrivial topological properties. Among the most intriguing is the state that arises at the Landau level filling factor of 5/2. Theoretical calculations suggest several possibilities for the 5/2 ground state and associated topological order, but distinguishing among them experimentally is tricky. Dutta *et al.* developed a method for doing so by interfacing a region in the 5/2 state with a region at an integer filling, and the measurements provided support for the particle-hole Pfaffian order. The technique can be used for the investigation of other exotic states in the quantum Hall setting. —JS

Science, abg6116, this issue p. 193

MOLECULAR MAGNETS
Magnetic effects of lanthanide bonding

Lanthanide coordination compounds have attracted attention for their persistent magnetic properties near liquid nitrogen temperature, well above alternative molecular magnets. Gould *et al.* report that introducing metal-metal bonding can enhance coercivity. Reduction of iodide-bridged terbium or dysprosium dimers resulted in a single electron bond between the metals, which enforced alignment of the other valence electrons. The resultant coercive fields exceeded 14 tesla below 50 and 60 kelvin for the terbium and dysprosium compounds, respectively. —JSY

Science, abl5470, this issue p. 198

MARTIAN GEOLOGY
Abiotic formation of organic molecules

Mars rovers have found complex organic molecules in the ancient rocks exposed on the planet's surface and methane in the modern atmosphere. It is unclear what processes produced these organics, with proposals including both biotic and abiotic sources. Steele *et al.* analyzed the nanoscale mineralogy of the Mars meteorite ALH 84001 and found evidence of organic synthesis driven by serpentinization and carbonation reactions that occurred during the aqueous alteration of basalt rock by hydrothermal fluids. The results demonstrate that abiotic production of organic molecules operated on Mars 4 billion years ago. —KTS

Science, abg7905, this issue p. 172

IMMUNOTHERAPY
Being a pro has its benefits

A key to tumor immunotherapy is revitalizing intratumoral immune cells. Cytokine therapy can rejuvenate these cells but causes immense toxicity, limiting the clinical efficacy of these treatments for cancer. Xue *et al.* formulated a shielded form of interleukin (IL)–12 called pro–IL-12 that was preferentially taken up by tumors where matrix metalloproteinases unshielded the cytokine. Systemic pro–IL-12 induced profound antitumor efficacy in multiple mouse tumor models, limiting toxicity compared with recombinant IL-12 treatment. Crucially, pro–IL-12 improved the efficacy of targeted inhibitors and immune checkpoint blockade in various tumor models without furthering

toxicity. Thus, pro–IL-12 is a safe therapy that could have a large clinical impact for patients with cancer. —DAE

Sci. Immunol. **6**, eabi6899 (2021).

CELL BIOLOGY
Ancient origin of cell death

Gasdermins are cell death proteins in mammals that form membrane pores in response to pathogen infection. Johnson *et al.* report that diverse bacteria encode structural and functional homologs of mammalian gasdermins. Like their mammalian counterparts, bacterial gasdermins are activated by caspase-like proteases, oligomerize into large membrane pores, and defend against pathogen—in this case, bacteriophage—infection. Proteolytic activation occurs through the release of a short inhibitory peptide, and many bacterial gasdermins are lipidated to facilitate membrane pore formation. Pyroptotic cell death, a central component of mammalian innate immunity, thus has a shared origin with an ancient antibacteriophage defense system. —SMH

Science, abj8432, this issue p. 221

PLANT SCIENCE
mRNA migration through plasmodesmata

In plants, certain transcription factors are produced in one cell but transported, sometimes as messenger RNA (mRNA), through plasmodesmata, channels between neighboring plant cells, where they act. This system helps to manage stem cell development. Kitagawa *et al.* have identified part of the machinery that manages this cell-to-cell transport. Transport of the mRNA encoding the KNOTTED1 homeobox transcription factor depends on Ribosomal RNA-Processing Protein 44 (AtRRP44A), a subunit of the RNA exosome. —PJH

Science, abm0840, this issue p. 177

2D MATERIALS
Following a crossover

Superfluidity in fermionic systems occurs through the pairing of fermions into bosons, which can undergo condensation. Depending on the strength of the interactions between fermions, the pairs range from large and overlapping to tightly bound. The crossover between these two limits has been explored in ultracold Fermi gases. Liu *et al.* observed the crossover in an electronic system consisting of two layers of graphene separated by an insulating barrier and placed in a magnetic field. In this two-dimensional system, the pairs were excitons formed from an electron in one layer and a hole in the other. The researchers used magnetic field and layer separation to tune the interactions and detected the signatures of superfluidity through transport measurements. —JS

Science, abg1110, this issue p. 205
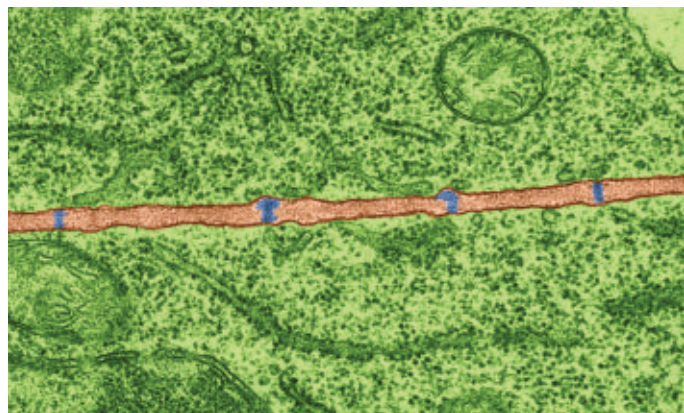
Plasmodesmata, seen here in blue in a false-color transmission electron micrograph, mediate the transmission of molecules as large as mRNA between plant cells.

IN OTHER JOURNALS
Edited by **Caroline Ash** *and* **Jesse Smith**

EVOLUTION
Dining out on transparency

Several butterflies mimic the coloration of poisonous species without being toxic themselves. Mimics benefit from predator avoidance without the cost of carrying the requisite metabolic machinery. Pinna *et al.* observed that some toxic butterflies with transparent wing patches have mimics too. The question is, what is being mimicked: camouflage, toxic potential, both, or some other advantage? Close examination of the diversity of microstructures of these patches indicated how they are precisely fine-tuned to bird predator perception and revealed convergent structural evolution among mimicking species. Transparent mimics avoid being spotted by hungry birds, but when they are, they are thought to be unpalatable because of their wing patterns and thus are rejected for lunch. —CA

eLife **10**, e69080 (2021).

Despite their transparency offering camouflage, glasswing butterflies may still mimic the wing patterns of toxic relatives to provide additional protection.

AQUATIC ECOLOGY
Wastewater degrades food webs

Urban wastewater has profound effects on aquatic ecosystems downstream, adding nutrients along with harmful pollutants. Mor *et al.* examined these effects across entire food webs, comparing bottom-up (increased primary production due to added nutrients) and top-down (loss of predators) effects. Observations from 10 streams showed that stream reaches exposed to effluent had higher primary production but lower energy transfer to consumers compared with paired upstream controls. They also found reduced species richness and narrower trophic niches in predators downstream from effluent. Thus, wastewater pollution has both direct (abiotic) and indirect (biotic) impacts on aquatic communities. —BEL

Ecology 10.1002/ecy.3587 (2021).

NUCLEAR PHYSICS
A proton-shedding nucleus

A stable atomic nucleus contains a careful balance of the number of protons and neutrons. For unstable nuclei that have a pronounced excess of protons, a rare radioactive process called proton emission, in which one or more protons are ejected from the nucleus, is the preferred road to stability. Although one to three proton emissions have been observed in several nuclides, only one nuclide has been observed to decay through a four-proton emission. Jin *et al.* now add ^{18}Mg to that short list. The researchers created the previously unobserved nuclide ^{18}Mg by smashing a beam of ^{20}Mg against a target and then detected ^{18}Mg through its decay to ^{14}O and four protons. —JS

Phys. Rev. Lett. **127**, 262502 (2021).

changes. Using optogenetic stimulation of GABAergic neurons of the hypothalamus, Yamagata *et al.* found that this area affects not only major state shifts but also within-state changes of sleep and wake intensity. The hypothalamus thus not only participates in switching global brain states but also regulates within-state intensity. —PRS

Proc. Natl. Acad. Sci. U.S.A. **118**, e2101580118 (2021).

STRUCTURAL BIOLOGY
Natural solution for air-water interfaces

Although advances in instrumentation and data processing have made cryo–electron microscopy a routine, high-resolution method for macromolecular structure determination, experiments are still often undermined by physical issues such as protein denaturation at the air-water interface or preferred orientation of particles. Fan *et al.* found that an amphipathic protein film can block the air-water interface and change the surface properties of the sample support. The authors demonstrate that this method leads to thin ice layers and improved behavior of proteins that have been challenging to work with previously. —MAF

Nat. Commun. **12**, 7257 (2021).

WEATHER SYSTEMS
Patterns of drying

Ridges of high atmospheric pressure can trap heat in a region for long periods and contribute to the occurrence of droughts. Are these types of highs responsible for the drying trend in the western United States that lasted from 1980 to 2018? Zhang *et al.* identified five distinct weather patterns that occur in that region and conclude that it was a decrease in the frequency of low-pressure troughs, which are associated with cooler and wetter conditions, rather than more high-pressure ridges, that led to the observed drying. This trend is likely a consequence of human activity. —HJS

Geophys. Res. Lett. 10.1029/2021GL097089 (2021).

AGING
Genetics shapes drug's effect

The beneficial effects of the anti-diabetes drug metformin appear to extend to heart health, cognitive function, and even increased health span and life span in some animal models. Onken *et al.* explored how metformin's effects varied in species and strains of worms in the genus *Caenorhabditis*, the genetic diversity of which approximates that seen between mice and humans. This could be informative for anticipated clinical trials of metformin's effects on aging in humans. However, variability was rather high. Some strains had increased life span, but in some it decreased. A measure of mobility, taken as an indicator of health, revealed that in some genetic backgrounds, health span could be increased without an increase in median or maximal life span. —LBR

Aging Cell 10.1111/acel.13488 (2021).

NEUROSCIENCE
The dance of sleep

Because of the presence or absence of consciousness, sleep and wakefulness are often thought to be two distinct global states. These brain states are characterized by unique intrinsic activity patterns. Sleep has a self-organized regimen characterized by changes in arrangement, as well as depth or intensity. It has long been believed that the preoptic area of the hypothalamus is the regulator of brain state

Sleep and wakefulness are different brain states regulated by the hypothalamus.

CELL BIOLOGY
Animal-free research products

Cell culture is used throughout biological research, but how cells are cultured may influence how they respond to treatments. Cell culture experiments for applied research, such as preclinical studies in human cells and testing pharmaceuticals and chemicals, may be affected by an ingredient used in cell culture media: fetal bovine serum (FBS). Although there are many potential sources of irreproducibility in research, the use of animal-derived materials such as FBS, which are inherently variable, may lead to alterations in the morphology, growth, viability, and responses of cells in culture. How FBS is sourced also raises ethical concerns. In a Perspective, van der Valk discusses the possible issues with using FBS in cell culture and the progress with synthetic media as an alternative. —GKA

Science, abm1317, this issue p. 143

CANCER
Targeting cyclin-dependent kinases

Cyclin-dependent kinases (CDKs), in complex with their cyclin partners, modulate the transition through phases of the cell division cycle. Cyclin D–CDK complexes are important in cancer progression, especially for certain types of breast cancer. Fassl *et al.* discuss advances in understanding the biology of cyclin D–CDK complexes that have led to new concepts about how drugs that target these complexes induce cancer cell cytostasis. The authors also suggest possible combinations to add to the types of cancer that can be treated and discuss progress in overcoming resistance to cyclin D–CDK inhibitors and their possible application to diseases beyond cancer. —GKA

Science, abc1495, this issue p. 158

CELL BIOLOGY
Mitochondria shed their SPOTs

Outer mitochondrial membrane (OMM) function is essential for cellular health. How mitochondria respond to naturally occurring OMM stress is unknown. Li *et al.* show that, upon infection with the human parasite *Toxoplasma gondii*, mitochondria shed large structures positive for OMM (SPOTs). SPOT formation required the parasite effector TgMAF1 and its interaction with the host mitochondrial receptor TOM70 and translocase SAM50. TOM70-dependent SPOT formation mediated a depletion of mitochondrial proteins and optimal parasite growth. SPOT-like structures also formed after OMM perturbations independently of infection. Thus, membrane remodeling is a feature of cellular responses to OMM stress that *Toxoplasma* hijacks during infection. —SMH

Science, abi4343, this issue p. 159

CORONAVIRUS
Persuading people to mask

Even in places where it is obligatory, people tend to optimistically overstate their compliance for mask wearing. How then can we persuade more of the population at large to act for the greater good? Abaluck *et al.* undertook a large, cluster-randomized trial in Bangladesh involving hundreds of thousands of people (although mostly men) over a 2-month period. Colored masks of various construction were handed out free of charge, accompanied by a range of mask-wearing promotional activities inspired by marketing research. Using a grassroots network of volunteers to help conduct the study and gather data, the authors discovered that mask wearing averaged 13.3% in villages where no interventions took place but increased to 42.3% in those where in-person interventions were introduced. Villages where in-person reinforcement of mask wearing occurred also showed a reduction in reporting COVID-like illness, particularly in high-risk individuals. —CA

Science, abi9069, this issue p. 160

CORONAVIRUS
Preparing antiviral defenses

Antiviral drugs are an important tool in the battle against COVID-19. Both remdesivir and molnupiravir, which target the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) polymerase, were first developed against other RNA viruses. This highlights the importance of broad-spectrum antivirals that can be rapidly deployed against related emerging pathogens. Sourimant *et al.* used respiratory syncytial virus (RSV) as a primary indication in identifying further drugs that target the polymerase enzyme of RNA viruses. The authors explored derivatives of molnupiravir and identified 4' fluorouridine (EIDD-2749) as an inhibitor of the polymerase of RSV and SARS-CoV-2. This drug can be delivered orally and was effective against RSV in mice and SARS-CoV-2 in ferrets. —VV

Science, abj5508, this issue p. 161

NEURODEGENERATION
Hi-res view of human Aβ42 filaments

Alzheimer's disease is characterized by a loss of memory and other cognitive functions and the filamentous assembly of Aβ and tau in the brain. The assembly of Aβ peptides into filaments that end at residue 42 is a central event. Yang *et al.* used cryo–electron microscopy to determine the structures of Aβ42 filaments from human brain (see the Perspective by Willem and Fändrich). They identified two types of related S-shaped filaments, each consisting of two identical protofilaments. These structures will inform the development of better in vitro and animal models, inhibitors of Aβ42 assembly, and imaging agents with increased specificity and sensitivity. —SMH

Science, abm7285, this issue p. 167; see also abn5428, p. 147

PLANETARY SCIENCE
Terapascal iron-melting temperature

The pressure and temperature conditions at which iron melts are important for terrestrial planets because they determine the size of the liquid metal core, an important factor for understanding the potential for generating a radiation-shielding magnetic field. Kraus *et al.* used laser-driven shock to determine the iron-melt curve up to a pressure of 1000 gigapascals (see the Perspective by Zhang and Lin). This value is about three times that of the Earth's inner core boundary. The authors found that the liquid metal core lasted the longest for Earth-like planets four to six times larger in mass than the Earth. —BG

Science, abm1472, this issue p. 202; see also abn2051, p. 146

PHYSICS
Gravitational interference

The Aharonov-Bohm effect is a quantum mechanical effect in which a magnetic field affects the phase of an electron wave as it propagates along a wire. Atom interferometry exploits the wave characteristic of atoms to measure tiny differences in phase as they take different paths through the arms of an interferometer. Overstreet *et al.* split a cloud of cold rubidium atoms into two atomic wave packets about 25 centimeters apart and subjected one of the wave packets to

gravitational interaction with a large mass (see the Perspective by Roura). The authors state that the observed phase shift is consistent with a gravitational Aharonov-Bohm effect. —ISO

Science, abl7152, this issue p. 226; see also abm6854, p. 142

CORONAVIRUS
Immune imprinting

For severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), immune responses to heterologous variants are influenced by a person's infection history. Healthcare workers (HCWs) may be exposed to several doses and types of antigens, either by natural infection or by vaccination. Reynolds *et al*. studied a cohort of UK HCWs followed since March 2020. The immunological profiles of these people depended on how often the subject had encountered antigen and which variant was involved. Vaccine responses after infection were found to be less effective if the infection involved heterologous spike from a variant virus. Unfortunately, the N501Y spike mutation, found in many variants, seems to induce the regulatory T cell transcription factor FOXP3, indicating that the virus could subvert effective T cell function. Changes to antibody binding between variants also means that serology data using the Wuhan Hu-1 S1 receptor-binding domain sequence may not be a reliable measure of protection. —CA

Science, abm0811, this issue p. 183

ECOLOGY
Seed dispersal in decline

Most plant species depend on animals to disperse their seeds, but this vital function is threatened by the declines in animal populations, limiting the potential for plants to adapt to climate change by shifting their ranges. Using data from more than 400 networks of seed dispersal interactions, Fricke *et al*. quantified the changes in seed disposal function brought about globally by defaunation. Their analyses indicate that past defaunation has severely reduced long-distance seed dispersal, cutting by more than half the number of seeds dispersed far enough keep up with range shifts caused by climate change. In addition, their approach enables the prediction of seed dispersal interactions using species traits and an estimation of how these interactions translate into ecosystem functioning, thus informing ecological forecasting and the consequences of animal declines. —AMS

Science, abk3510, this issue p. 210

ATHEROSCLEROSIS
Sniffing out atherosclerosis

Olfactory receptors are best known for their presence in the nose and their role in detecting smells, but they are also present in other tissues and perform additional biological functions. For example, vascular macrophages involved in the pathogenesis of atherosclerosis express multiple subtypes of olfactory receptors. Orecchioni *et al*. focused on olfactory receptor 2, a receptor for the compound octanal, and identified its contribution to atherosclerosis pathogenesis and the formation of atherosclerotic plaques (see the Perspective by Rayner and Rasheed). The authors show that most of the octanal was not directly derived from the diet, but rather was generated as a by-product of lipid peroxidation, suggesting a potential pathway for intervention. —YN

Science, abg3067, this issue p. 214; see also abn4708, p. 145

REGENERATIVE MEDICINE
Exciting scaffolds regenerate cartilage

Osteoarthritis is a common condition, but curative therapies are limited. Liu *et al.* developed a biodegradable scaffold using poly(L-lactic acid) (PLLA) nanofibers that, when placed under applied force, generated a piezoelectric charge, leading to improved chondrogenesis in vitro. Rabbits with osteochondral defects in the medial femoral condyle that were treated with the PLLA scaffold demonstrated improved cartilage regeneration and subchondral bone regeneration after 1 or 2 months of exercise to generate piezoelectric charge from the joint loads. Although further work is needed to optimize the scaffold microstructure and components, this study provides evidence that biodegradable piezoelectric scaffolds can use joint-loading exercise to treat osteoarthritis. —MN

Sci. Transl. Med. **13**, eabi7282 (2022).

CELL BIOLOGY
Redox control of GPCR desensitization

Activated G protein–coupled receptors (GPCRs) undergo internalization and desensitization. Nishiyama *et al*. identified a mechanism through which electrophilic compounds modified the purinergic receptor $P2Y_6R$, leading to ubiquitylation, receptor internalization, and proteasomal degradation. Mice expressing a form of $P2Y_6R$ that could not be electrophilically modified were more susceptible to experimentally induced colitis. The positions of these cysteine and lysine residues were conserved in several other GPCRs, suggesting that this mechanism of desensitization may be shared by other GPCRs. —AMV

Sci. Signal. **15**, eabj0644 (2022).

CANCER

CDK4 and CDK6 kinases: From basic science to cancer therapy

Anne Fassl, Yan Geng, Piotr Sicinski*

BACKGROUND: Cyclins and cyclin-dependent kinases (CDKs) drive cell division. Of particular importance to the cancer field are D-cyclins, which activate CDK4 and CDK6. In normal cells, the activity of cyclin D–CDK4/6 is controlled by the extracellular pro-proliferative or inhibitory signals. By contrast, in many cancers, cyclin D–CDK4/6 kinases are hyperactivated and become independent of mitogenic stimulation, thereby driving uncontrolled tumor cell proliferation. Mouse genetic experiments established that cyclin D–CDK4/6 kinases are essential for growth of many tumor types, and

they represent potential therapeutic targets. Genetic and cell culture studies documented the dependence of breast cancer cells on CDK4/6. Chemical CDK4/6 inhibitors were synthesized and tested in preclinical studies. Introduction of these compounds to the clinic represented a breakthrough in breast cancer treatment and will likely have a major impact on the treatment of many other tumor types.

ADVANCES: Small-molecule CDK4/6 inhibitors (palbociclib, ribociclib, abemaciclib) showed impressive results in clinical trials for patients

with hormone receptor–positive breast cancers. Addition of CDK4/6 inhibitors to standard endocrine therapy substantially extended median progression-free survival and prolonged median overall survival. Consequently, all three CDK4/6 inhibitors have been approved for treatment of women with advanced or metastatic hormone receptor–positive breast cancers. In the past few years, the renewed interest in CDK4/6 biology has yielded several surprising discoveries. The emerging concept is that CDK4/6 kinases regulate a much wider set of cellular functions than anticipated. Consequently, CDK4/6 inhibitors, beyond inhibiting tumor cell proliferation, affect tumor cells and the tumor environment through mechanisms that are only beginning to be elucidated. For example, inhibition of CDK4/6 affects antitumor immunity acting both on tumor cells and on the host immune system. CDK4/6 inhibitors were shown to enhance the efficacy of immune checkpoint blockade in preclinical mouse cancer models. These new concepts are now being tested in clinical trials.

OUTLOOK: Palbociclib, ribociclib, and abemaciclib are being tested in more than 300 clinical trials for more than 50 tumor types. These trials evaluate CDK4/6 inhibitors in combination with a wide range of therapeutic compounds that target other cancer-relevant pathways. Several other combination treatments were shown to be efficacious in preclinical studies and will enter clinical trials soon. Another CDK4/6 inhibitor, trilaciclib, is being tested for its ability to shield normal cells of the host from cytotoxic effects of chemotherapy. New CDK4/6 inhibitors have been developed and are being assessed in preclinical and clinical trials. The major impediment in the therapeutic use of CDK4/6 inhibitors is that patients who initially respond to treatment often develop resistance and eventually succumb to the disease. Moreover, a substantial fraction of tumors show preexisting, intrinsic resistance to CDK4/6 inhibitors. One of the main challenges will be to elucidate the full range of resistance mechanisms. Even with the current, limited knowledge, one can envisage the principles of new, improved approaches to overcome known resistance mechanisms. Another largely unexplored area for future study is the possible involvement of CDK4/6 in other pathologic states beyond cancer. This will be the subject of intense studies, and it may extend the utility of CDK4/6 inhibitors to the treatment of other diseases. ∎
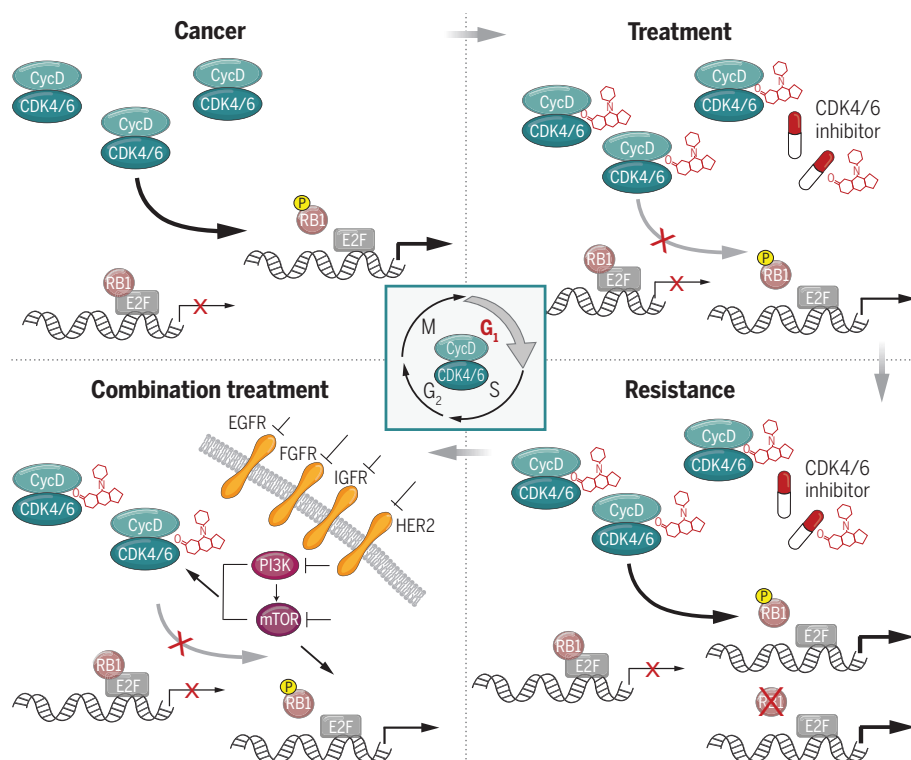
Targeting cyclin D–CDK4/6 for cancer treatment. D-cyclins (CycD) activate CDK4 and CDK6 in G_1 phase of the cell cycle and promote cell cycle progression by phosphorylating the retinoblastoma protein RB1. RB1 inhibits E2F transcription factors; phosphorylation of RB1 activates E2F-driven transcription. In many cancers, CycD-CDK4/6 is constitutively activated and drives uncontrolled cell proliferation. The development of small-molecule CDK4/6 inhibitors provided a therapeutic tool to repress constitutive CycD-CDK4/6 activity and to inhibit cancer cell proliferation. As with several targeted therapies, tumors eventually develop resistance and resume cell proliferation despite CDK4/6 inhibition. New combination treatments, involving CDK4/6 inhibitors plus inhibition of other pathways, are being tested in the clinic to delay or overcome the resistance.

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abc1495

CANCER

CDK4 and CDK6 kinases: From basic science to cancer therapy

Anne Fassl, Yan Geng, Piotr Sicinski*

Cyclin-dependent kinases 4 and 6 (CDK4 and CDK6) and their activating partners, D-type cyclins, link the extracellular environment with the core cell cycle machinery. Constitutive activation of cyclin D–CDK4/6 represents the driving force of tumorigenesis in several cancer types. Small-molecule inhibitors of CDK4/6 have been used with great success in the treatment of hormone receptor–positive breast cancers and are in clinical trials for many other tumor types. Unexpectedly, recent work indicates that inhibition of CDK4/6 affects a wide range of cellular functions such as tumor cell metabolism and antitumor immunity. We discuss how recent advances in understanding CDK4/6 biology are opening new avenues for the future use of cyclin D–CDK4/6 inhibitors in cancer treatment.

yclin D1, the activator of CDK4 and CDK6, was discovered in the early 1990s (*1*, *2*). The role of cyclin D1 in oncogenesis was already evident at the time of its cloning, as it was also identified as the protein product of the *PRAD1* oncogene, which is rearranged and overexpressed in parathyroid adenomas (*3*), and of the *BCL1* oncogene, which is rearranged in B-lymphocytic malignancies (*4*). Subsequently, the remaining two D-type cyclins, D2 and D3, were discovered on the basis of their homology to cyclin D1 (*1*).

Cyclins serve as regulatory subunits of cyclin-dependent kinases (CDKs) (*5*). Shortly after the discovery of D-cyclins, CDK4 and CDK6 were identified as their kinase partners (*6*). Mouse gene knockout studies revealed that CDK4 and CDK6 play redundant roles in development, and combined ablation of CDK4 and CDK6 was found to result in embryonic lethality (*7*). The essentially identical phenotype was seen in cyclin D–knockout mice, thereby confirming the role of D-cyclins as CDK4/6 activators in vivo (*8*). Surprisingly, these analyses revealed that many normal nontransformed mammalian cell types can proliferate without any cyclin D–CDK4/6 activity (*7*, *8*).

CDK4 and CDK6 are expressed at constant levels throughout the cell cycle. By contrast, D-cyclins are labile proteins that are transcriptionally induced upon stimulation of cells with growth factors. For this reason, D-cyclins are regarded as links between the cellular environment and the cell cycle machinery (*6*).

Cell cycle inhibitors play an important role in regulating the activity of cyclin D–CDK4/6 (Fig. 1). The INK inhibitors (p16[INK4A], p15[INK4B], p18[INK4C], p19[INK4D]) bind to CDK4 or CDK6 and prevent their interaction with D-type cyclins, thereby inhibiting cyclin D–CDK4/6 kinase activity. By contrast, KIP/CIP inhibitors (p27[KIP1], p57[KIP2], p21[CIP1]), which inhibit the activity of CDK2-containing complexes, serve as assembly factors for cyclin D–CDK4/6 (*6*, *9*). This was demonstrated by the observation that mouse fibroblasts devoid of p27[KIP1] and p21[CIP1] fail to assemble cyclin D–CDK4/6 complexes (*10*).

p27[KIP1] can bind cyclin D–CDK4/6 in an inhibitory or noninhibitory mode, depending on p27[KIP1] phosphorylation status. Cyclin D–p27[KIP1]-CDK4/6 complexes are catalytically inactive unless p27[KIP1] is phosphorylated on Tyr[88] and Tyr[89] (*11*). Two molecular mechanisms may explain this switch. First, Tyr[88]/Tyr[89] phosphorylation may dislodge the helix of p27[KIP1] from the CDK active site and allow adenosine triphosphate (ATP) binding (*12*). Second, the presence of tyrosine-unphosphorylated p27[KIP1] within the cyclin D–CDK4 complex prevents the activating phosphorylation of CDK4's T-loop by the CDK-activating kinase (CAK) (*12*). Brk has been identified as a physiological kinase of p27[KIP1] (*13*); Abl and Lyn can phosphorylate p27[KIP1] in vitro, but their in vivo importance remains unclear (*11*, *14*).

The activity of cyclin D–CDK4/6 is also regulated by proteolysis. Cyclin D1 is an unstable

---

Box 1. Clinical use of CDK4/6 inhibitors.

Palbociclib

Approved by FDA in 2016, in combination with fulvestrant for the treatment of hormone receptor–positive, HER2-negative (HR+/HER2−) advanced or metastatic breast cancer in women with disease progression following endocrine therapy. Approved in 2017 for the treatment of HR+/HER2− advanced or metastatic breast cancer in combination with an aromatase inhibitor as initial endocrine-based therapy in postmenopausal women.

Palbociclib is administered at a dose of 125 mg (given orally) daily for 3 weeks followed by 1 week off, or 200 mg daily for 2 weeks followed by 1 week off. The rate-limiting toxicities are neutropenia, thrombocytopenia, and anemia.

Ribociclib

Approved by FDA in 2017, in combination with an aromatase inhibitor as initial endocrine-based therapy for the treatment of postmenopausal women with HR+/HER2− advanced or metastatic breast cancer. In 2018, the FDA expanded the indication for ribociclib in combination with an aromatase inhibitor for pre/perimenopausal women with HR+/HER2− advanced or metastatic breast cancer, as initial endocrine-based therapy. FDA also approved ribociclib in combination with fulvestrant for postmenopausal women with HR+/HER2− advanced or metastatic breast cancer, as initial endocrine-based therapy or following disease progression on endocrine therapy.

Ribociclib is administered at a dose of 600 mg (given orally) daily for 3 weeks followed by 1 week off. The main toxicities are neutropenia and thrombocytopenia.

Abemaciclib

Approved by FDA in 2017, in combination with fulvestrant for women with HR+/HER2− advanced or metastatic breast cancer with disease progression following endocrine therapy. In addition, abemaciclib was approved as monotherapy for women and men with HR+/HER2− advanced or metastatic breast cancer with disease progression following endocrine therapy and prior chemotherapy in the metastatic setting. Approved by FDA in 2018 in combination with an aromatase inhibitor as initial endocrine-based therapy for postmenopausal women with HR+/HER2− advanced or metastatic breast cancer. Approved by FDA in 2021 for adjuvant treatment of early-stage HR+/HER2− breast cancer in combination with endocrine therapy.

Abemaciclib is administered at a dose of 200 mg (given orally) every 12 hours. The dose-limiting toxicity is fatigue. Neutropenia is also observed but is not rate-limiting. Other severe side effects include diarrhea and nausea.

Department of Cancer Biology, Dana-Farber Cancer Institute, Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02215, USA.
*Corresponding author. Email: peter_sicinski@dfci.harvard.edu

protein with a half-life of less than 30 min. At the end of G_1 phase, cyclin D1 is phosphorylated at Thr^{286} by GSK3β (15). This facilitates association of cyclin D1 with the nuclear exportin CRM1 and promotes export of cyclin D1 from the nucleus to the cytoplasm (16). Subsequently, phosphorylated cyclin D1 becomes polyubiquitinated by E3 ubiquitin ligases, thereby targeting it for proteasomal degradation. Several substrate receptors of E3 ubiquitin ligases have been implicated in recognizing phosphorylated cyclin D1, including F-box proteins FBXO4 (along with αB crystallin), FBXO31, FBXW8, β-TrCP1/2, and SKP2 (17). The anaphase-promoting complex/cyclosome (APC/C) was also proposed to target cyclin D1 while F-box proteins FBXL2 and FBXL8 target cyclins D2 and D3 (17, 18). Surprisingly, the level and stability of cyclin D1 was unaffected by depletion of several of these proteins, indicating that some other E3 plays a rate-limiting role in cyclin D1 degradation (19). Indeed, recent studies reported that D-cyclins are ubiquitinated and targeted for proteasomal degradation by the E3 ubiquitin ligase CRL4, which uses AMBRA1 protein as its substrate receptor (20–22).

Cyclin D–CDK4/6 in cancer

Genomic aberrations of the cyclin D1 gene (*CCND1*) represent frequent events in different tumor types. The t(11;14)(q13;q32) translocation juxtaposing *CCND1* with the immunoglobulin heavy-chain (*IGH*) locus represents the characteristic feature of mantle-cell lymphoma and is frequently observed in multiple myeloma or plasma cell leukemia (23, 24). Amplification of *CCND1* is seen in many other malignancies—for example, in 13 to 20% of breast cancers (23, 24), more than 40% of head and neck squamous cell carcinomas, and more than 30% of esophageal squamous cell carcinomas (23). A higher proportion of cancers (e.g., up to 50% of mammary carcinomas) overexpress cyclin D1 protein (24). Also, cyclins D2 and D3, CDK4, and CDK6 are overexpressed in various tumor types (5, 9). Cyclin D–CDK4/6 can also be hyperactivated through other mechanisms such as deletion or inactivation of INK inhibitors, most frequently $p16^{INK4A}$ (5, 9, 23). Altogether, a very large number of human tumors contain lesions that hyperactivate cyclin D–CDK4/6 (5).

An oncogenic role for cyclin D–CDK4/6 has been supported by mouse cancer models. For example, targeted overexpression of cyclin D1 in mammary glands of transgenic mice led to the development of mammary carcinomas (25). Also, overexpression of cyclin D2, D3, or CDK4, or loss of $p16^{INK4a}$ resulted in tumor formation (9).

Conversely, genetic ablation of D-cyclins, CDK4, or CDK6 decreased tumor sensitivity (9). For instance, *Ccnd1-* or *Cdk4*-null mice, or
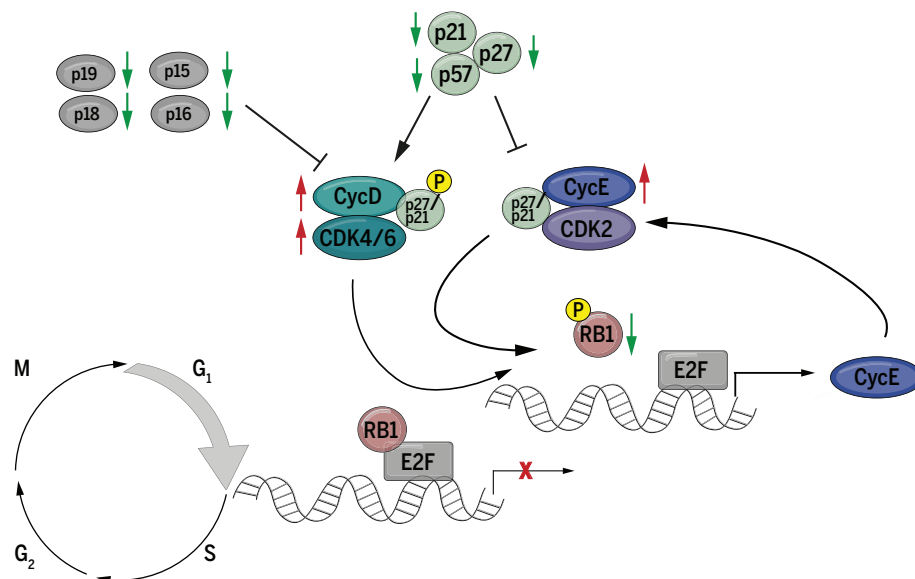
knock-in mice expressing kinase-inactive cyclin D1–CDK4/6, were resistant to develop human epidermal growth factor receptor 2 (HER2)–driven mammary carcinomas (26–29). An acute, global shutdown of cyclin D1 in mice bearing HER2-driven tumors arrested tumor growth and triggered tumor-specific senescence while having no obvious impact on normal tissues (30). Likewise, an acute ablation of CDK4 arrested tumor cell proliferation and triggered tumor cell senescence in a KRAS-driven non–small-cell lung cancer (NSCLC) mouse model (31). These observations indicated that CDK4 and CDK6 might represent excellent therapeutic targets in cancer treatment.

CDK4/6 functions in cell proliferation and oncogenesis

The best-documented function of cyclin D–CDK4/6 in driving cell proliferation is phosphorylation of the retinoblastoma protein, RB1, and RB-like proteins, RBL1 and RBL2 (5, 6) (Fig. 1). Unphosphorylated RB1 binds and inactivates or represses E2F transcription factors. According to the prevailing model, phosphorylation of RB1 by cyclin D–CDK4/6 partially inactivates RB1, leading to release of E2Fs and up-regulation of E2F-transcriptional targets, including cyclin E. Cyclin E forms a complex with its kinase partner, CDK2, and completes full RB1 phosphorylation, leading

to activation of the E2F transcriptional program and facilitating S-phase entry (5, 6). In normal, nontransformed cells, the activity of cyclin D–CDK4/6 is tightly regulated by the extracellular mitogenic milieu. This links inactivation of RB1 with mitogenic signals. In cancer cells carrying activating lesions in cyclin D–CDK4/6, the kinase is constitutively active, thereby decoupling cell division from proliferative and inhibitory signals (5).

This model has been questioned by the demonstration that RB1 exists in a monophosphorylated state throughout G_1 phase and becomes inactivated in late G_1 by cyclin E–CDK2, which "hyperphosphorylates" RB1 on multiple residues (32). However, recent single-cell analyses revealed that cyclin D–CDK4/6 activity is required for the hyperphosphorylation of RB1 throughout G_1, whereas cyclin E/A–CDK maintains RB1 hyperphosphorylation in S phase (33). Moreover, phosphorylation of RB1 by cyclin D–CDK4/6 was shown to be required for normal cell cycle progression (34).

In addition to this kinase-dependent mechanism, up-regulation of D-cyclin expression and formation of cyclin D–CDK4/6 complexes lead to redistribution of KIP/CIP inhibitors from cyclin E–CDK2 complexes (which are inhibited by these proteins) to cyclin D–CDK4/6 (which use them as assembly factors), thereby activating the kinase activity of cyclin E–CDK2 (6). Cyclin E–CDK2 in turn phosphorylates RB1

Fig. 1. Molecular events governing progression through the G_1 phase of the cell cycle. The mammalian cell cycle can be divided into G_1, S (DNA synthesis), G_2, and M (mitosis) phases. During G_1 phase, cyclin D (CycD)–CDK4/6 kinases together with cyclin E (CycE)–CDK2 phosphorylate the retinoblastoma protein RB1. This activates the E2F transcriptional program and allows entry of cells into S phase. Members of the INK family of inhibitors ($p16^{INK4A}$, $p15^{INK4B}$, $p18^{INK4C}$, and $p19^{INK4D}$) inhibit cyclin D–CDK4/6; KIP/CIP proteins ($p21^{CIP1}$, $p27^{KIP1}$, and $p57^{KIP2}$) inhibit cyclin E–CDK2. Cyclin D–CDK4/6 complexes use $p27^{KIP1}$ and $p21^{CIP1}$ as "assembly factors" and sequester them away from cyclin E–CDK2, thereby activating CDK2. Proteins that are frequently lost or down-regulated in cancers are marked with green arrows, overexpressed proteins with red arrows.

and other cellular proteins and promotes cell cycle progression.

Cyclin D1–CDK4/6 directly phosphorylates, stabilizes, and activates the transcription factor FOXM1. This promotes cell cycle progression and protects cancer cells from entering senescence (35). Cyclin D–CDK4 also phosphorylates and inactivates SMAD3, which mediates transforming growth factor–β (TGF-β) antiproliferative response. CDK4/6-dependent phosphorylation of SMAD3 inhibits its transcriptional activity and disables the ability of TGF-β to induce cell cycle arrest (36). FZR1/CDH1, an adaptor protein of the APC complex, is another phosphorylation substrate of CDK4. Depletion of CDH1 in human cancer cells partially rescued the proliferative block upon CDK4/6 inhibition, and it cooperated with RB1 depletion in restoring full proliferation (37).

Cyclin D–CDK4/6 also phosphorylates and inactivates TSC2, a negative regulator of mTORC1, thereby resulting in mTORC1 activation. Conversely, inhibition of CDK4/6 led to decreased mTORC1 activity and reduced protein synthesis in cells representing different human tumor types. It was proposed that through TSC2 phosphorylation, activation of cyclin D–CDK4/6 couples cell growth with cell division (38). Consistent with this, the antiproliferative effect of CDK4/6 inhibition was reduced in cells lacking TSC2 (38).

MEP50, a co-regulatory factor of protein arginine-methyltransferase 5 (PRMT5), is phosphorylated by cyclin D1–CDK4. Through this mechanism, cyclin D1–CDK4/6 increases the catalytic activity of PRMT5/MEP50 (39). It was proposed that deregulation of cyclin D1–CDK4 kinase in tumor cells, by increasing PRMT5/MEP50 activity, reduces the expression of CUL4, a component of the E3 ubiquitin-ligase complex, and stabilizes CUL4 targets such as CDT1 (39). In addition, by stimulating PRMT5/MEP50-dependent arginine methylation of p53, cyclin D–CDK4/6 suppresses the expression of key antiproliferative and pro-apoptotic p53 target genes (40). Another study proposed that PRMT5 regulates splicing of the transcript encoding MDM4, a negative regulator of p53. CDK4/6 inhibition reduced PRMT5 activity and altered the pre-mRNA splicing of MDM4, leading to decreased levels of MDM4 protein and resulting in p53 activation. This, in turn, up-regulated the expression of a p53 target, p21[CIP1], that blocks cell cycle progression (41).

During oncogenic transformation of hematopoietic cells, chromatin-bound CDK6 phosphorylates the transcription factors NFY and SP1 and induces the expression of p53 antagonists such as PRMT5, PPM1D, and MDM4 (42). Also, in acute myeloid leukemia cells expressing constitutively activated FLT3, CDK6 binds the promoter region of the FLT3 gene as well as the promoter of PIM1 pro-oncogenic kinase and stimulates their expression. Treatment of FLT3-mutant leukemic cells with a CDK4/6 inhibitor decreased FLT3 and PIM1 expression and triggered cell cycle arrest and apoptosis (43). The relevance of these various

Table 1. Currently available CDK4/6 inhibitors. This table lists major inhibitors of CDK4 and CDK6, half-maximal inhibitory concentration (IC$_{50}$) for different cyclin D–CDK4/6 complexes (if known), other known targets, and the stage of clinical development. K_i, inhibitory constant.

Name of compound	IC$_{50}$	Other known targets	Stage of clinical development
Palbociclib (PD-0332991)	D1-CDK4, 11 nM; D2-CDK6, 15 nM; D3-CDK4, 9 nM		FDA-approved for HR$^+$/HER2$^-$ advanced breast cancer in combination with endocrine therapy; phase 2/3 trials for several other tumor types
Ribociclib (LEE011)	D1-CDK4, 10 nM; D3-CDK6, 39 nM		FDA-approved for HR$^+$/HER2$^-$ advanced breast cancer in combination with endocrine therapy; phase 2/3 trials for several other tumor types
Abemaciclib (LY2835219)	D1-CDK4, 0.6 to 2 nM; D3-CDK6, 8 nM	Cyclin T1–CDK9, PIM1, HIPK2, CDKL5, CAMK2A, CAMK2D, CAMK2G, GSK3α/β, and (at higher doses) cyclin E/A–CDK2 and cyclin B–CDK1	FDA-approved for early (adjuvant) and advanced HR$^+$/HER2$^-$ breast cancer in combination with endocrine therapy; FDA-approved as monotherapy in advanced HR$^+$/HER2$^-$ breast cancer; phase 2/3 trials for several other tumor types
Trilaciclib (G1T28)	D1-CDK4, 1 nM; D3-CDK6, 4 nM		FDA-approved for small-cell lung cancer to reduce chemotherapy-induced bone marrow suppression; phase 2/3 trials for other solid tumors
Lerociclib (G1T38)	D1-CDK4, 1 nM; D3-CDK6, 2 nM		Phase 1/2 trials for HR$^+$/HER2$^-$ advanced breast cancer and EGFR-mutant non–small-cell lung cancer
SHR6390	CDK4, 12 nM; CDK6, 10 nM		Phase 1/2/3 trials for HR$^+$/HER2$^-$ advanced breast cancer and other solid tumors
PF-06873600	CDK4, 0.13 nM (K_i), CDK6, 0.16 nM (K_i)	CDK2, 0.09 nM (K_i)	Phase 2 trials for HR$^+$/HER2$^-$ advanced breast cancer and other solid tumors
FCN-437	D1-CDK4, 3.3 nM; D3-CDK6, 13.7 nM		Phase 1/2 trials for HR$^+$/HER2$^-$ advanced breast cancer and other solid tumors
Birociclib (XZP-3287)	Not reported		Phase 1/2 trials for HR$^+$/HER2$^-$ advanced breast cancer and other solid tumors
HS-10342	Not reported		Phase 1/2 trials for HR$^+$/HER2$^-$ advanced breast cancer and other solid tumors
CS3002	Not reported		Phase 1 trial for solid tumors

Mechanism of action of CDK4/6 inhibitors

Three small-molecule CDK4/6 inhibitors have been extensively characterized in preclinical studies: palbociclib and ribociclib, which are highly specific CDK4/6 inhibitors, and abemaciclib, which inhibits CDK4/6 and other kinases (Table 1). It has been assumed that these compounds act in vivo by directly inhibiting cyclin D–CDK4/6 (9). This simple model has been recently questioned by observations that palbociclib inhibits only cyclin D–CDK4/6 dimers, but not trimeric cyclin D–CDK4/6-p27[KIP1] (44). However, it is unlikely that substantial amounts of cyclin D–CDK4 dimers ever exist in cells, because nearly all cyclin D–CDK4 in vivo is thought to be complexed with KIP/CIP proteins (11, 14, 44). Palbociclib also binds monomeric CDK4 (44). Surprisingly, treatment of cancer cells with palbociclib for 48 hours failed to inhibit CDK4 kinase, despite cell cycle arrest, but it inhibited CDK2 (44). Hence, palbociclib might prevent the formation of active CDK4-containing complexes (through binding to CDK4) and indirectly inhibit CDK2 by liberating KIP/CIP inhibitors. This model needs to be reconciled with several observations. First, treatment of cells with CDK4/6 inhibitors results in a rapid decrease of RB1 phosphorylation on cyclin D–CDK4/6-dependent sites, indicating an acute inhibition of CDK4/6 (45–47). Moreover, CDK4/6 immunoprecipitated from cells can be inhibited by palbociclib (48) and p21[CIP]-associated cyclin CDK4/6 kinase is also inhibited by treatment of cells with palbociclib (49). Lastly, CDK2 is dispensable for proliferation of several cancer cell lines (50, 51), hence the indirect inhibition of CDK2 alone is unlikely to be responsible for cell cycle arrest.

Palbociclib, ribociclib, and abemaciclib were shown to block binding of CDK4 and CDK6 to CDC37, the kinase-targeting subunit of HSP90, thereby preventing access of CDK4/6 to the HSP90-chaperone system (52). Because the HSP90-CDC37 complex stabilizes several kinases (53), these observations suggest that CDK4/6 inhibitors, by disrupting the interaction between CDC37 and CDK4 or CDK6, might promote degradation of CDK4 and CDK6. However, depletion of CDK4/6 is typically not observed upon treatment with CDK4/6 inhibitors (54). More studies are needed to resolve these conflicting reports and to establish how CDK4/6 inhibitors affect the cell cycle machinery in cancer cells.

Validation of CDK4/6 inhibitors as anticancer agents

Consistent with the notion that RB1 represents the major rate-limiting substrate of cyclin D–CDK4/6 in cell cycle progression (55–57),

palbociclib, ribociclib, and abemaciclib were shown to block proliferation of several RB1-positive cancer cell lines, but not cell lines that have lost RB1 expression (46, 58, 59). Breast cancer cell lines representing the luminal, estrogen receptor–positive (ER[+]) subtype were shown to be most susceptible to cell proliferation arrest upon palbociclib treatment (45). Palbociclib, ribociclib, abemaciclib, and another CDK4/6 inhibitor, lerociclib, were demonstrated to display potent antitumor activity in xenografts of several tumor types, including breast cancers (46, 60–62). Palbociclib and abemaciclib cross the blood-brain barrier and inhibit growth of intracranial glioblastoma (GBM) xenografts, with abemaciclib being more efficient in reaching the brain (63, 64). Recently, additional CDK4/6 inhibitors were shown to exert therapeutic effects in mouse xenograft models of various cancer types, including SHR6390 (65), FCN-437 (66), and compound 11 (67); the latter two were reported to cross the blood-brain barrier. In most in vivo studies, the therapeutic effect was dependent on expression of intact RB1 protein in tumor cells (46, 63). However, antitumor effects of palbociclib were also reported in bladder cancer xenografts independently of RB1 status; this was attributed to decreased phosphorylation of FOXM1 (68).

Tumor cell senescence upon CDK4/6 inhibition

In addition to blocking cell proliferation, inhibition of CDK4/6 can also trigger tumor cell senescence (63), which depends on RB1 and FOXM1 (35, 54). The role of RB1 in enforcing cellular senescence is well established (69). In addition, cyclin D–CDK4/6 phosphorylates and activates FOXM1, which has anti-senescence activity (35, 70). Senescence represents a preferred therapeutic outcome to cell cycle arrest, as it may lead to a durable inhibition of tumor growth.

It is not clear what determines the extent of senescence upon treatment of cancer cells with CDK4/6 inhibitors. A recent study showed that inhibition of CDK4/6 leads to an RB1-dependent increase in reactive oxygen species (ROS) levels, resulting in activation of autophagy, which mitigates the senescence of breast cancer cells in vitro and in vivo (71). Co-treatment with palbociclib plus autophagy inhibitors strongly augmented the ability of CDK4/6 inhibitors to induce tumor cell senescence and led to sustained inhibition of cancer cell proliferation in vitro and of xenograft growth in vivo (71). Decreased mTOR signaling after long-term CDK4/6 inhibition was shown to be essential for the induction of senescence in melanoma cells, and activation of mTORC1 overrode palbociclib-induced senescence (72). Others postulated that expression of the chromatin-remodeling enzyme

ATRX and degradation of MDM2 determines the choice between quiescence and senescence upon CDK4/6 inhibition (73). Inhibition of CDK4 causes dissociation of the deubiquitinase HAUSP/USP7 from MDM2, thereby driving autoubiquitination and proteolytic degradation of MDM2, which in turn promotes senescence. This mechanism requires ATRX, which suggests that expression of ATRX can be used to predict the senescence response (73). Two additional proteins that play a role in this process are PDLIM7 and type II cadherin CDH18. Expression of CDH18 correlated with a sustained response to palbociclib in a phase 2 trial for patients with liposarcoma (74).

Markers predicting response to CDK4/6 inhibition

Only tumors with intact RB1 respond to CDK4/6 inhibitor treatment by undergoing cell cycle arrest or senescence (9, 58). In addition, "D-cyclin activating features" (CCND1 translocation, CCND2 or CCND3 amplification, loss of the CCND1-3 3′-untranslated region, and deletion of FBXO31 encoding an F-box protein implicated in cyclin D1 degradation) were shown to confer a strong response to abemaciclib in cancer cell lines (58). Moreover, co-deletion of CDKN2A and CDKN2C (encoding p16[INK4A]/p19[ARF] and p18[INK4C], respectively) confers palbociclib sensitivity in glioblastoma (75). Thr[172] phosphorylation of CDK4 and Tyr[88] phosphorylation of p27[KIP1] (both associated with active cyclin D–CDK4) correlate with sensitivity of breast cancer cell lines or tumor explants to palbociclib (76, 77). Surprisingly, in PALOMA-1, PALOMA-2, and PALOMA-3 trials (78–80), and in another independent large-scale study (81), CCND1 gene amplification or elevated levels of cyclin D1 mRNA or protein were not predictive of palbociclib efficacy. Conversely, overexpression of CDK4, CDK6, or cyclin E1 is associated with resistance of tumors to CDK4/6 inhibitors (see below).

Synergy of CDK4/6 inhibitors with other compounds

Several preclinical studies have documented the additive or synergistic effects of combining CDK4/6 inhibitors with inhibitors of the receptor tyrosine kinases as well as phosphoinositide 3-kinase (PI3K), RAF, or MEK (Table 2). This synergism might be because these pathways impinge on the cell cycle machinery through cyclin D–CDK4/6 (82–86). In some cases, the effect was seen in the presence of specific genetic lesions, such as EGFR, BRAF[V600E], KRAS, and PIK3CA mutations (59, 87–89) (Table 2). When comparing different dosing regimens, continuous treatment with a MEK inhibitor with intermittent palbociclib resulted in more complete tumor responses than other combination schedules (90). Treatment with CDK4/6 inhibitors sensitized cancer cells to ionizing

radiation (*63*) or cisplatin (*68*). The synergism with platinum-based chemotherapy was attributed to the observation that upon this treatment, CDK6 phosphorylates and stabilizes the FOXO3 transcription factor, thereby promoting tumor cell survival. Consequently, inhibition of CDK6 increases platinum sensitivity by enhancing tumor cell death (*91*).

In several instances, co-treatment with CDK4/6 inhibitors prevented the development of resistance to other compounds or inhibited the proliferation of resistant tumor cells. Co-treatment of melanoma patient-derived xenografts (PDXs) with ribociclib plus the RAF inhibitor encorafenib delayed or prevented development of encorafenib resistance (*92*). PDXs that acquired encorafenib resistance remained sensitive to the combination of encorafenib plus ribociclib (*59*). Treatment of BRAF[V600E]-mutant melanoma xenografts with palbociclib plus the BRAF[V600E] inhibitor PLX4720 prevented development of resistance (*89*). BRAF[V600E]-mutant melanoma cell lines that acquired resistance to the BRAF[V600E] inhibitor vemurafenib remained sensitive to palbociclib or abemaciclib, and xenografts underwent senescence and tumor regression upon CDK4/6 inhibition (*72*, *93*). Treatment of

ALK-mutant, ALK kinase inhibitor–resistant neuroblastoma xenografts with palbociclib restored the sensitivity to these compounds (*94*). A combination of PI3K and CDK4/6 inhibitors overcame the intrinsic and acquired resistance of breast cancers to PI3K inhibitors and resulted in regression of *PIK3CA*-mutant xenografts (*88*).

Up-regulation of cyclin D1 expression was shown to mediate acquired resistance of HER2[+] tumors to anti-HER2 therapies in a mouse breast cancer model (*95*). Treatment of mice bearing trastuzumab-resistant tumors or PDXs of resistant HER2[+] mammary carcinomas with abemaciclib restored the sensitivity of tumors to HER2 inhibitors and inhibited tumor cell proliferation. Moreover, in the case of treatment-naïve tumors, co-administration of abemaciclib significantly delayed the development of resistance to anti-HER2 therapies (*95*).

Several anticancer treatments, such as chemotherapy, target dividing cells. Because CDK4/6 inhibitors block tumor cell proliferation, they might impede the effects of chemotherapy. Indeed, several reports have documented that co-administration of CDK4/6 inhibitors antagonized the antitumor effects of compounds that act during S phase (doxorubicin, gemcitabine,

methotrexate, mercaptopurine) or mitosis (taxanes) (*96*, *97*). However, some authors reported synergistic effects (*98*, *99*), although the molecular underpinnings are unclear.

A recent report documented that administration of CDK4/6 inhibitors prior to taxanes inhibited tumor cell proliferation and impeded the effect of taxanes (*100*). By contrast, administration of taxanes first (or other chemotherapeutic compounds that act on mitotic cells or cells undergoing DNA synthesis), followed by CDK4/6 inhibitors, had a strong synergistic effect. The authors showed that by repressing the E2F-dependent transcriptional program, CDK4/6 inhibitors impaired the expression of genes required for DNA-damage repair via homologous recombination. Because treatment of cancer cells with chemotherapy triggers DNA damage, the impairment of DNA-damage repair induced cytotoxicity, thereby explaining the synergistic effect (*100*).

Cells with impaired homologous recombination rely on poly-(ADP-ribose) polymerase (PARP) for double-stranded DNA-damage repair, which renders them sensitive to PARP inhibition. Indeed, a strong synergistic effect has been demonstrated between CDK4/6

Table 2. Combination treatments that demonstrated synergy with CDK4/6 inhibitors in preclinical studies. TNBC, triple-negative breast cancer; AR, androgen receptor; ER[+], estrogen receptor–positive; T-ALL, T cell acute lymphoblastic leukemia; HER2[+], human epidermal growth factor receptor 2–positive; PI3K, phosphoinositide 3-kinase; EGFR, epidermal growth factor receptor; IGF1R, insulin-like growth factor 1 receptor, InsR, insulin receptor.

CDK4/6 inhibitor	Synergistic target	Inhibitor	Disease
Palbociclib	PI3K	Taselisib, pictilisib	PIK3CA mutant TNBC
	AR	Enzalutamide	Androgen receptor–positive TNBC
	EGFR	Erlotinib	TNBC, esophageal squamous cell carcinoma
	RAF	PLX4720	BRAF-V600E mutant melanoma
	MEK	Trametinib	KRAS mutant colorectal cancer
	MEK	PD0325901 (mirdametinib)	KRAS or BRAFV600E mutant colorectal cancer
	MEK	MEK162 (binimetinib)	KRAS mutant colorectal cancer
	MEK	AZD6244 (selumetinib)	Pancreatic ductal adenocarcinoma
	PI3K/mTOR	BEZ235 (dactolisib), AZD0855, GDC0980 (apitolisib)	Pancreatic ductal adenocarcinoma
	IGF1R/InsR	BMS-754807	Pancreatic ductal adenocarcinoma
	mTOR	Temsirolimus	Pancreatic ductal adenocarcinoma
	mTOR	AZD2014 (vistusertib)	ER[+] breast cancer
	mTOR	MLN0128 (sapanisertib)	Intrahepatic cholangiocarcinoma
	mTOR	Everolimus	Melanoma, glioblastoma
Ribociclib	PI3K	GDC-0941 (pictilisib), BYL719 (alpelisib)	PIK3CA mutant breast cancer
	PDK1	GSK2334470	ER[+] breast cancer
	EGFR	Nazartinib	EGFR-mutant lung cancer
	RAF	Encorafenib	BRAF-V600E mutant melanoma
	mTOR	Everolimus	T-ALL
	Inflammation	Glucocorticoid dexamethasone	T-ALL
	γ-Secretase	Compound E	T-ALL
Abemaciclib	HER2	Trastuzumab	HER2[+] breast cancer
	EGFR and HER2	Lapatinib	HER2[+] breast cancer
	RAF	LY3009120, vemurafenib	KRAS mutant lung or colorectal cancer, NRAS or BRAF-V600E mutant melanoma
		Temozolomide (alkylating agent)	Glioblastoma

inhibitors and PARP inhibitors in PDX-derived cell lines (*100*). Such synergy was also reported for ovarian cancer cells (*101*). Another study found that inhibition of CDK4/6 resulted in down-regulation of PARP levels (*102*).

Protection against chemotherapy-induced toxicity

Administration of palbociclib to mice induced reversible quiescence in hematopoietic stem/progenitor cells (HSPCs). This effect protected mice from myelosuppression after total-body irradiation. Moreover, treatment of tumor-bearing mice with CDK4/6 inhibitors together with irradiation mitigated radiation-induced toxicity without compromising the therapeutic effect (*103*). Co-administration of a CDK4/6 inhibitor, trilaciclib, with cytotoxic chemotherapy (5-FU, etoposide) protected animals from chemotherapy-induced exhaustion of HSPCs, myelosuppression, and apoptosis of bone marrow (*60*, *104*). These observations led to phase 2 clinical trial, which evaluated the effects of trilaciclib administered prior to etoposide and carboplatin for treatment of small-cell lung cancer. Trilaciclib improved myelopreservation while having no adverse effect on antitumor efficacy (*105*). A similar phase 2 clinical trial investigating trilaciclib in combination with gemcitabine and carboplatin chemotherapy in patients with metastatic triple-negative breast cancer (TNBC) did not observe a significant difference in myelosuppression. However, this study demonstrated an overall survival benefit of the combination therapy (*106*, *107*).

Metabolic function of CDK4/6 in cancer cells

The role of CDK4/6 in tumor metabolism is only starting to be appreciated (Fig. 2A). Treatment of pancreatic cancer cells with CDK4/6 inhibitors was shown to induce tumor cell metabolic reprogramming (*108*). CDK4/6 inhibition increased the numbers of mitochondria and lysosomes, activated mTOR, and increased the rate of oxidative phosphorylation, likely through an RB1-dependent mechanism (*108*). Combined inhibition of CDK4/6 and mTOR strongly suppressed tumor cell proliferation (*108*). Moreover, CDK4/6 can phosphorylate and inactivate TFEB, the master regulator of lysosomogenesis, and through this mechanism reduce lysosomal numbers. Conversely, CDK4/6 inhibition activated TFEB and increased the number of lysosomes (*109*). Another mechanism linking CDK4/6 and lysosomes was provided by the observation that treatment of TNBC cells with CDK4/6 inhibitors decreased mTORC1 activity and impaired the recruitment of mTORC1 to lysosomes (*110*). Consistent with the idea that mTORC1 inhibits lysosomal biogenesis, CDK4/6 inhibition increased the number of lysosomes in tumor cells. Because an increased lysosomal biomass
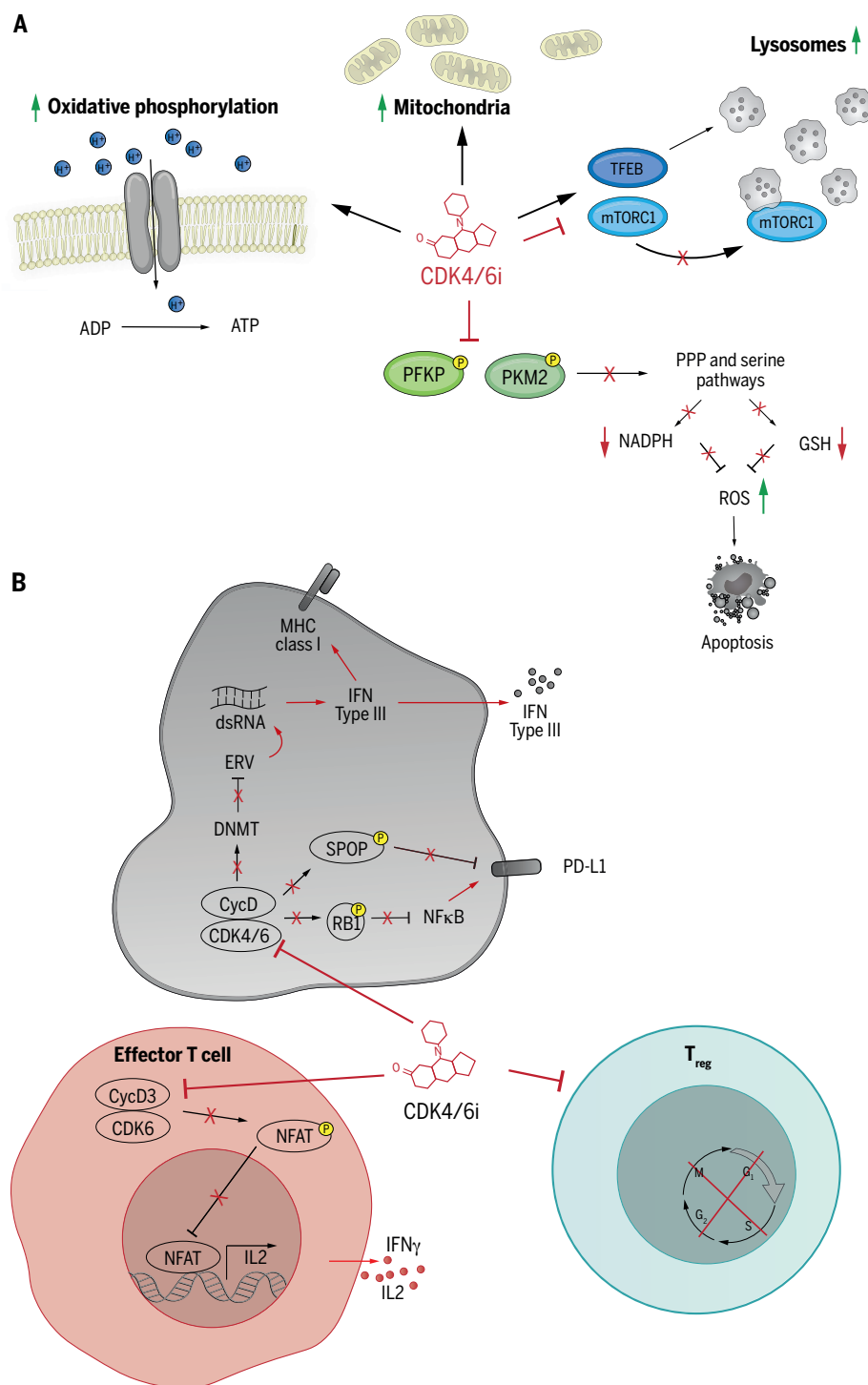

Fig. 2. CDK4 and CDK6: More than cell cycle kinases. Although the role of CDK4 and CDK6 in cell cycle progression has been well documented, both kinases regulate several other functions that are only now starting to be unraveled. (**A**) Inhibition of CDK4/6 (CDK4/6i) affects lysosome and mitochondrial numbers as well as oxidative phosphorylation. Cyclin D3–CDK6 phosphorylates glycolytic enzymes 6-phosphofructokinase (PFKP) and pyruvate kinase M2 (PKM2), thereby controlling ROS levels via the pentose phosphate (PPP) and serine synthesis pathways. (**B**) Inhibition of CDK4/6 affects antitumor immunity, acting both within cancer cells and on the immune system of the host. In tumor cells, inhibition of CDK4/6 impedes expression of an E2F target, DNA methyltransferase (DNMT). DNMT inhibition reduces methylation of endogenous retroviral genes (ERV) and increases intracellular levels of double-stranded RNA (dsRNA) (*114*). In effector T cells, inhibition of CDK4/6 stimulates NFAT transcriptional activity and enhances secretion of IFN-γ and interleukin 2 (IL-2) (*115*).

underlies some cases of CDK4/6 inhibitor resistance (see below) (*111*), stimulation of lysosomogenesis by CDK4/6 inhibitors might limit their clinical efficacy by inducing resistance.

Lastly, CDK4/6 inhibition impaired lysosomal function and the autophagic flux in cancer cells. It was argued that this lysosomal dysfunction was responsible for the senescent phenotype in CDK4/6 inhibitor–treated cells (*110*). Because lysosomes are essential for autophagy, the authors co-treated TNBC xenografts with abemaciclib plus an AMPK activator, A769662 (which induces autophagy), and found that this led to cancer cell death and subsequent regression of tumors (*110*).

Cyclin D3–CDK6 phosphorylates and inhibits two rate-limiting glycolytic enzymes, 6-phosphofructokinase and pyruvate kinase M2. This redirects glycolytic intermediates into the pentose phosphate pathway (PPP) and serine synthesis pathway. Through this mechanism, cyclin D3–CDK6 promotes the production of reduced nicotinamide adenine dinucleotide phosphate (NADPH) and reduced glutathione (GSH) and helps to neutralize ROS (*112*). Treatment of tumors expressing high levels of cyclin D3–CDK6 (such as leukemias) with CDK4/6 inhibitors reduced the PPP- and serine-synthesis pathway flow, thereby depleting the antioxidants NADPH and GSH. This increased ROS levels and triggered tumor cell apoptosis (*112*).

Another link between cyclin D–CDK4/6 in metabolism and cancer was provided by the observation that livers of obese/diabetic mice up-regulate cyclin D1 expression (*113*). Treatment of these mice with an antidiabetic compound, metformin, reduced liver cyclin D1 levels and largely protected mice against development of hepatocellular carcinoma. Also, genetic ablation of cyclin D1 protected obese/diabetic mice from liver cancer, and administration of palbociclib inhibited liver cancer progression. These treatments had no effect on tumors in nonobese animals (*113*). These observations raise the possibility of using antidiabetic compounds with CDK4/6 inhibitors for treatment of liver cancers in obese patients.

CDK4/6 inhibitors and antitumor immune responses

Several recent reports have started to unravel how inhibition of CDK4/6 influences antitumor immune responses, acting both on tumor cells as well as on the tumor immune environment (Fig. 2B). Treatment of breast cancer–bearing mice or breast cancer cells with abemaciclib activated expression of endogenous retroviral elements in tumor cells, thereby increasing the levels of double-stranded RNA. This, in turn, stimulated production of type III interferons and increased presentation of tumor antigens. Hence, CDK4/6 inhibitors, by inducing viral gene expression, trigger anti-

viral immune responses that help to eliminate the tumor (*114*).

Inhibition of CDK4/6 also affects the immune system by impeding the proliferation of $CD4^+FOXP3^+$ regulatory T cells (T_{regs}), which normally inhibit the antitumor response. Because cytotoxic $CD8^+$ T cells are less affected by CDK4/6 inhibition, abemaciclib treatment decreases the $T_{reg}/CD8^+$ ratio of intratumoral T cells and facilitates tumor cell killing by cytotoxic $CD8^+$ T cells (*114*).

Inhibition of CDK4/6 also resulted in activation of T cells through derepression of NFAT signaling. NFAT4 (and possibly other NFATs) are phosphorylated by cyclin D3–CDK6 (*115*). Inhibition of CDK4/6 decreased phosphorylation of NFATs, resulting in their nuclear translocation and enhanced transcriptional activity. This caused up-regulation of NFAT targets, resulting in T cell activation, which enhanced the antitumor immune response. In addition, CDK4/6 inhibitors increased the infiltration of effector T cells into tumors, likely because of elevated levels of chemokines CXCL9 and CXCL10 after CDK4/6 inhibitor treatment (*115*). Abemaciclib treatment also induced inflammatory and activated T cell phenotypes in tumors and up-regulated the expression of immune checkpoint proteins CD137, PD-L1, and TIM-3 on $CD4^+$ and $CD8^+$ cells (*116*).

CDK4/6 inhibition also caused up-regulation of PD-L1 protein expression in tumor cells (*117*, *118*). This effect was shown to be independent of RB1 status in the tumor. Mechanistically, CDK4/6 phosphorylates and stabilizes SPOP, which promotes PD-L1 polyubiquitination and degradation (*118*). Cyclin D–CDK4 also represses expression of PD-L1 through RB1. Specifically, cyclin D–CDK4/6-mediated phosphorylation of RB1 on S249/T252 promotes binding of RB1 to NF-κB protein p65, and this represses the expression of a subset NF-κB–regulated genes, including PD-L1 (*119*).

These observations prompted tests of the efficacy of combining CDK4/6 inhibitors with antibodies that elicit immune checkpoint blockade. Indeed, treatment of mice bearing autochthonous breast cancers, or cancer allografts, with CDK4/6 inhibitors together with anti-PD-1/PD-L1 antibodies enhanced the efficacy of immune checkpoint blockade and led to complete tumor regression in a high proportion of animals (*114*, *115*, *118*). Conversely, activation of the cyclin D–CDK4 pathway by genomic lesions in human melanomas correlated with resistance to anti–PD-1 therapy (*117*).

Some authors did not observe synergy when abemaciclib was administered concurrently with immune checkpoint inhibitors in allograft tumor models (*116*, *120*). However, a strong synergistic antitumor effect was detected when abemaciclib was administered first (and continued) and anti–PD-L1 antibody was administered later. The combined treatment induced

immunological memory, as mice that underwent tumor regression were resistant to rechallenge with the same tumor (*116*). Abemaciclib plus anti–PD-L1 treatment increased infiltration of $CD4^+$ and $CD8^+$ T cells into tumors, and increased the expression of major histocompatibility complex class I (MHC-I) and MHC-II on tumor cells and on macrophages and MHC-I on dendritic cells (*116*). In the case of anti–CTLA-4 plus anti–PD-1 treatment in melanoma allograft model, the synergistic effect was observed when immune checkpoint inhibitor treatment was started first, followed by abemaciclib (*120*).

The synergistic antitumor effect of PI3K and CDK4/6 inhibitors in TNBC is mediated, in part, by enhancement of tumor immunogenicity (*121*). Combined treatment of TNBC cells with ribociclib plus the PI3K inhibitor apelisib synergistically up-regulated the expression of immune-related pathways in tumor cells, including proteins involved in antigen presentation. Co-treatment of tumor-bearing mice also decreased proliferation of $CD4^+FOXP3^+$ T_{reg} cells, increased activation of intratumoral $CD4^+$ and $CD8^+$ T cells, increased the frequency of tumor-infiltrating NKT cells, and decreased the numbers of intratumoral immunosuppressive myeloid-derived suppressor cells. Moreover, combined treatment strongly augmented the response to immune checkpoint therapy with PD-1 and CTLA-4 antibodies (*121*).

Single-cell RNA sequencing of human melanomas identified an immune resistance program expressed by tumor cells that correlates with T cell exclusion from the tumor mass and immune evasion by tumor cells. The program can predict the response of tumors to immune checkpoint inhibitors. Treatment of human melanoma cells with abemaciclib repressed this program in an RB1-dependent fashion (*120*).

Together, these findings indicate that CDK4/6 inhibitors may convert immunologically "cold" tumors into "hot" ones. The most pressing issue is to validate these findings in a clinical setting. The utility of combining CDK4/6 inhibitors with PD-1 or PD-L1 antibodies is currently being evaluated in several clinical trials. Note that the effects of CDK4/6 inhibition on the immune system of the host are independent of tumor cell RB1 status, raising the possibility of using CDK4/6 inhibitors to also boost the immune response against RB1-negative tumors.

CDK4/6 inhibitors in clinical trials

Table 3 summarizes major clinical trials with CDK4/6 inhibitors. Given early preclinical data indicating that breast cancers—in particular, the hormone receptor–positive ones—are very sensitive to CDK4/6 inhibition (as discussed above), many clinical trials have focused on this cancer type. Most studies have evaluated CDK4/6 inhibitors administered together with anti-estrogens (the aromatase inhibitors

Table 3. Major past clinical trials with CDK4/6 inhibitors. ER⁺, estrogen receptor–positive; HER2⁻, human epidermal growth factor receptor 2–negative; HR⁺, hormone receptor–positive; PFS, progression-free survival. FGFR2, fibroblast growth factor receptor 2; ERBB3, receptor tyrosine-protein kinase erbB-3.

CDK4/6 inhibitor	Trial name	Trial details	Treatment	Patients	Outcome	Ref.	Other outcomes
Palbociclib	PALOMA-1	Randomized phase 2	Aromatase inhibitor letrozole alone (standard of care) versus letrozole plus palbociclib	Postmenopausal women with advanced ER⁺/HER2⁻ breast cancer who had not received any systemic treatment for their advanced disease	Addition of palbociclib markedly increased median PFS from 10.2 months in the letrozole group to 20.2 months in the palbociclib plus letrozole group	(78)	On the basis of this result, palbociclib received a "Breakthrough Therapy" designation status from FDA and was granted accelerated approval, in combination with letrozole, for the treatment of ER⁺/HER2⁻ metastatic breast cancer
Palbociclib	PALOMA-2	Double-blind phase 3	Palbociclib plus letrozole as first-line therapy	Postmenopausal women with ER⁺/HER2⁻ breast cancer	Addition of palbociclib strongly increased median PFS: 14.5 months in the placebo-letrozole group versus 24.8 months in the palbociclib-letrozole group	(123)	Palbociclib was equally efficacious in patients with luminal A and B breast cancers, and there was no single biomarker associated with the lack of clinical benefit, except for RB1 loss; CDK4 amplification was associated with endocrine resistance, but this was mitigated by addition of palbociclib; tumors with high levels of FGFR2 and ERBB3 mRNA displayed greater PFS gain after addition of palbociclib (79)
Palbociclib	PALOMA-3	Randomized phase 3	Estrogen receptor antagonist fulvestrant plus placebo versus fulvestrant plus palbociclib	Women with HR⁺/HER2⁻ metastatic breast cancer that had progressed on previous endocrine therapy	The study demonstrated a substantial prolongation of median PFS in the palbociclib-treated group: 4.6 months in the placebo plus fulvestrant group versus 9.5 months in the palbociclib plus fulvestrant group; addition of palbociclib also extended median overall survival from 28.0 months (placebo-fulvestrant) to 34.9 months (palbociclib-fulvestrant); estimated rate of survival at 3 years was 41% versus 50%, respectively	(124, 125, 135)	
Palbociclib	NeoPalAna	Palbociclib in an neoadjuvant setting (i.e., prior to surgery)	Compared the effects of an aromatase inhibitor anastrozole versus palbociclib plus anastrozole on tumor cell proliferation	Women with newly diagnosed clinical stage II/III ER⁺/HER2⁻ breast cancer	Addition of palbociclib enhanced the antiproliferative effect of anastrozole	(161)	

continued on next page

CDK4/6 inhibitor	Trial name	Trial details	Treatment	Patients	Outcome	Ref.	Other outcomes
Palbociclib	PALLAS	Randomized phase 3	Palbociclib plus standard endocrine therapy versus endocrine therapy alone	Patients with early (stage 2 or 3), HR⁺/HER2⁻ breast cancer	Preliminary results indicate that the trial is unlikely to show a statistically significant improvement of invasive disease-free survival	(138)	
Palbociclib	PENELOPE-B		Palbociclib in patients with early breast cancer at high risk of recurrence		Ongoing		
Ribociclib	MONALEESA-2	Randomized phase 3	Ribociclib plus letrozole versus placebo plus letrozole	First-line treatment for postmenopausal women with HR⁺/HER2⁻ recurrent or metastatic breast cancer who had not received previous systemic therapy for advanced disease	At 18 months, PFS was 42.2% in the placebo-letrozole group and 63.0% in the ribociclib-letrozole group	(126)	
Ribociclib	MONALEESA-3	Phase 3	Ribociclib plus fulvestrant	Patients with advanced (metastatic or recurrent) HR⁺/HER2⁻ breast cancer who have either received no treatment for the advanced disease or previously received a single line of endocrine therapy for the advanced disease	Addition of ribociclib significantly extended median PFS, from 12.8 months (placebo-fulvestrant) to 20.5 months (ribociclib-fulvestrant); overall survival at 42 months was also extended from 45.9% (placebo-fulvestrant) to 57.8% (ribociclib-fulvestrant)	(127, 133)	
Ribociclib	MONALEESA-7	Phase 3 randomized, double-blind	Ribociclib versus placebo together with an anti-estrogen tamoxifen or an aromatase inhibitor (letrozole or anastrozole)	Premenopausal and perimenopausal women with HR⁺/HER2⁻ advanced breast cancer who had not received previous treatment with CDK4/6 inhibitors	Ribociclib significantly increased median PFS from 13.0 months in the placebo-endocrine therapy group to 23.8 months in the ribociclib-endocrine therapy group; overall survival was also strongly prolonged in the ribociclib group (estimated overall survival at 42 months was 46.0% for the placebo group and 70.2% in the ribociclib group)	(128, 132)	
Ribociclib	EarLEE-1	Phase 3 trial	Ribociclib in the treatment of early-stage, high-risk HR⁺/HER2 breast cancers		Ongoing		

continued on next page

CDK4/6 inhibitor	Trial name	Trial details	Treatment	Patients	Outcome	Ref.	Other outcomes
Abemaciclib	MONARCH 1	Phase 2 trial	Abemaciclib as a single agent	Women with HR$^+$/HER2$^-$ metastatic breast cancer who had progressed on or after prior endocrine therapy and had 1 or 2 chemotherapy regimens in the metastatic setting	Abemaciclib exhibited promising activity in these heavily pretreated patients with poor prognosis; median PFS was 6.0 months and overall survival 17.7 months	(136)	The most common adverse events were diarrhea, fatigue, and nausea (136)
Abemaciclib	MONARCH 2	Double-blind phase 3	Abemaciclib in combination with fulvestrant	Women with HR$^+$/HER2$^-$ breast cancer who had progressed while receiving endocrine therapy, or while receiving first-line endocrine therapy for metastatic disease	Addition of abemaciclib significantly increased PFS from 9.3 months in the placebo-fulvestrant to 16.4 in the abemaciclib-fulvestrant group; median overall survival was also extended from 37.3 months to 46.7 months	(129, 134)	
Abemaciclib	MONARCH 3	Randomized phase 3 double-blind	Abemaciclib plus an aromatase inhibitor (anastrozole or letrozole)	Postmenopausal women with advanced HR$^+$/HER2$^-$ breast cancer who had no prior systemic therapy in the advanced setting	Addition of abemaciclib prolonged PFS from 14.8 months (in the placebo-aromatase inhibitor group) to 28.2 months (abemaciclib-aromatase inhibitor group)	(130, 131)	
Abemaciclib	MonarchE	Phase 3 study	Endocrine with or without abemaciclib	Patients with HR$^+$/HER2$^-$ lymph node–positive, high-risk early breast cancer	Preliminary analysis indicates that addition of abemaciclib resulted in a significant improvement of invasive disease-free survival and of distant relapse-free survival	(137)	
Trilaciclib		Randomized phase 2 study	Chemotherapy alone (gemcitabine and carboplatin), versus concurrent administration of trilaciclib plus chemotherapy, versus administration of trilaciclib prior to chemotherapy (to mitigate the cytotoxic effect of chemotherapy on bone marrow)	Patients with recurrent or metastatic triple-negative breast cancer who had no more than two previous lines of chemotherapy	Addition of trilaciclib did not offer detectable myeloprotection, but resulted in increased overall survival (from 12.8 months in the chemotherapy-only group to 20.1 months in the concurrent trilaciclib and chemotherapy group and 17.8 months in trilaciclib before chemotherapy group)	(162)	The most common adverse events were neutropenia, thrombocytopenia, and anemia (162)

letrozole or anastrozole, or the estrogen receptor antagonist fulvestrant) for treatment of advanced/metastatic HR$^+$/HER2$^-$ breast cancers in postmenopausal women. Addition of CDK4/6 inhibitors significantly extended median progression-free survival ([78, 122–130]) and prolonged median overall survival ([131–134]). Moreover, abemaciclib has shown clinical activity when administered as a single agent ([135]). Consequently, palbociclib, ribociclib, and abemaciclib have been approved by the US Food and Drug Administration (FDA) for treatment of patients with advanced/metastatic HR$^+$/HER2$^-$ breast cancer (Box 1). A recent phase 3 clinical trial, MonarchE, evaluated abemaciclib plus standard endocrine therapy in treatment of patients with early-stage, high-risk, lymph node–positive HR$^+$/HER2$^-$ breast cancer. Addition of abemaciclib reduced the risk of breast cancer recurrence ([136]). This is in contrast to the similar PALLAS study reported this year, which found no benefit of adding palbociclib to endocrine therapy for women with early-stage breast cancer ([137]). Analysis of patient populations in these two trials may help to explain the different outcomes. It is also possible that the favorable outcome of the MonarchE study reflects a broader spectrum of kinases inhibited by abemaciclib. The utility of CDK4/6 inhibitors in early-stage breast cancer remains unclear and is being addressed in ongoing clinical trials (PALLAS, PENELOPE-B, EarLEE-1, MonarchE) ([138]). Currently, palbociclib is being used in 164 active or recruiting clinical trials, ribociclib in 69 trials, and abemaciclib in 98 trials for more than 50 tumor types ([139]). These trials evaluate combinations of CDK4/6 inhibitors with a wide range of compounds (Table 4). Trials with trilaciclib test the benefit of this compound in preserving bone marrow and the immune system.

Resistance to CDK4/6 inhibitors

Although CDK4/6 inhibitors represent very effective agents in cancer treatment, nearly all patients eventually develop resistance and succumb to the disease. Moreover, a substantial fraction of tumors show intrinsic resistance to treatment with CDK4/6 inhibitors (Fig. 3).

The best-documented mechanism of preexisting and acquired resistance is the loss of RB1 ([71, 81, 140]). Acquired RB1 loss has been detected in PDXs ([141]), in circulating tumor DNA (ctDNA) ([142, 143]), and in tumors from patients treated with CDK4/6 inhibitors ([144, 145]). However, RB1 mutations are likely subclonal and are seen in only 5 to 10% of patients ([143, 145]).

Increased expression of CDK6 was shown to underlie acquired resistance to CDK4/6 inhibitors. Amplification of the *CDK6* gene and the resulting overexpression of CDK6 protein were found in abemaciclib-resistant ER$^+$ breast cancer cells ([146]) and in ctDNA of patients with

ER$^+$ breast cancers that progressed during treatment with palbociclib plus endocrine therapy ([147]). Also, *CDK4* gene amplification conferred insensitivity to CDK4/6 inhibition in GBM and sarcomas ([148–150]), whereas overexpression of CDK4 protein was associated with resistance to endocrine therapy in HR$^+$ breast cancers ([79]).

Resistant breast cancer cells can also upregulate the expression of CDK6 through suppression of the TGF-β/SMAD4 pathway by the microRNA miR-432-5p. In this mechanism, exosomal expression of miR-432-5p mediates the transfer of the resistance phenotype between neighboring cell populations ([151]). Another mechanism of CDK6 up-regulation in ER$^+$ breast cancers is the loss of FAT1, which represses CDK6 expression via the Hippo pathway. Loss of FAT1 triggers up-regulation of CDK6 expression by the Hippo pathway effectors TAZ and YAP. Moreover, genomic alterations in other components of the Hippo pathway, although rare, are also associated with reduced sensitivity to CDK4/6 inhibitors ([81]).

Genetic lesions that activate pathways converging on D-type cyclins can cause resistance to CDK4/6 inhibitors. These include (i) *FGFR1/2* gene amplification or mutational activation, detected in ctDNA from patients with ER$^+$ breast cancers that progressed upon treatment with palbociclib plus endocrine therapy ([147]); (ii) hyperactivation of the MAPK pathway in resistant prostate adenocarcinoma cells, possibly due to increased production of EGF by cancer cells ([152]); and (iii) increased secretion of FGF in palbociclib-resistant KRAS-mutant NSCLC cells, which stimulates FGFR1 signaling in an autocrine or paracrine fashion, resulting in activation of ERK1/2 and mTOR as well as up-regulation of D-cyclin, CDK6, and cyclin E expression ([153]). Analyses of longitudinal tumor biopsies from a melanoma patient revealed an activating mutation in the *PIK3CA* gene that conferred resistance to ribociclib plus MEK inhibitor treatment ([154]). It is possible that these lesions elevate the cellular levels of active cyclin D–CDK4/6 complexes, thereby increasing the threshold for CDK4/6 inhibition.

Formation of a noncanonical cyclin D1–CDK2 complex was shown to represent another mechanism of acquired CDK4/6 inhibitor resistance. Such a complex was observed in palbociclib-treated ER$^+$ breast cancer cells and was implicated in overcoming palbociclib-induced cell cycle arrest ([141]). Also, depletion of AMBRA1 promoted the interaction of D-cyclins with CDK2, resulting in resistance to CDK4/6 inhibitors ([20, 22]); it remains to be seen whether this represents an intrinsic or acquired resistance mechanism in human tumors.

Genetic analyses revealed that activation of cyclin E can bypass the requirement for cyclin D–CDK4/6 in development and tumorigenesis

([155, 156]). Hence, it comes as no surprise that increased activity of cyclin E–CDK2 is responsible for a large proportion of intrinsic and acquired resistance to CDK4/6 inhibitors. Several different mechanisms can activate cyclin E–CDK2 kinase in resistant tumor cells: (i) Down-regulation of KIP/CIP inhibitors results in increased activity of cyclin E–CDK ([54, 157]). (ii) Loss of PTEN expression, which activates AKT signaling, leads to nuclear exclusion of p27^{KIP1}. This in turn prevents access of p27^{KIP1} to CDK2, resulting in increased CDK2 kinase activity ([144]). (iii) Activation of the PI3K/AKT pathway causes decreased levels of p21^{CIP1}. Co-treatment of melanoma PDXs with MDM2 inhibitors (which up-regulate p21^{CIP1} via p53) sensitized intrinsically resistant tumor cells to CDK4/6 inhibitors ([158]). (iv) Up-regulation of cyclin D1 levels triggers sequestration of KIP/CIP inhibitors from cyclin E–CDK2 to cyclin D–CDK4/6, thereby activating the former ([158]). (v) Amplification of the *CCNE1* gene and increased levels of cyclin E1 protein result in elevated activity of E-CDK2 kinase ([141]). (vi) mTOR signaling has been shown to up-regulate cyclin E1 (and D1) in KRAS-mutated pancreatic cancer cells; CDK2 activity was essential for CDK4/6 inhibitor resistance in this setting ([159]). (vii) Up-regulation of PDK1 results in activation of the AKT pathway, which increases the expression of cyclins E and A and activates CDK2 ([160]). (viii) In CDK4/6 inhibitor–resistant melanoma cells, high levels of RNA-binding protein FXR1 increase translation of the amino acid transporter SLC36A1. Up-regulation of SLC36A1 expression activates mTORC1, which in turn increases CDK2 expression ([161]). All these lesions are expected to allow cell proliferation, despite CDK4/6 inhibition, as a consequence of the activation of the downstream cell cycle kinase CDK2.

The role for cyclin E–CDK2 in CDK4/6 inhibitor resistance has been confirmed in clinical trials. In patients with advanced ER$^+$ breast cancer treated with palbociclib and letrozole or fulvestrant, the presence of proteolytically cleaved cytoplasmic cyclin E in tumor tissue conferred strongly shortened progression-free survival ([71]). Moreover, analyses of PALOMA-3 trial for patients with ER$^+$ breast cancers revealed lower efficacy of palbociclib plus fulvestrant in patients displaying high cyclin E mRNA levels in metastatic biopsies ([80]). Amplification of the *CCNE1* gene was detected in ctDNA of patients with ER$^+$ breast cancers that progressed on palbociclib plus endocrine therapy ([147]). Also, amplification of the *CCNE2* gene (encoding cyclin E2) was seen in a fraction of CDK4/6 inhibitor–resistant HR$^+$ mammary carcinomas ([145, 162]).

Collectively, these analyses indicate that resistant cells may become dependent on CDK2 for cell cycle progression. Indeed, depletion of CDK2 or inhibition of CDK2 kinase activity in

Table 4. Ongoing clinical trials testing new combinations with CDK4/6 inhibitors. HR$^+$, hormone receptor–positive; LHRH, luteinizing hormone–releasing hormone; ER$^+$, estrogen receptor–positive; PD-1, programmed cell death protein 1; PD-L1, programmed cell death 1 ligand 1; AR$^+$, androgen receptor–positive; TNBC, triple-negative breast cancer; EGFR, epidermal growth factor receptor; HER2$^+$, human epidermal growth factor receptor 2–positive; FGFR, fibroblast growth factor receptor; IGFR, insulin-like growth factor receptor; VEGF, vascular endothelial growth factor receptor; PI3K, phosphoinositide 3-kinase; NSCLC, non–small-cell lung cancer; ALL, acute lymphoblastic leukemia; SCLC, small-cell lung cancer.

Additional target	Inhibitor	Immune checkpoint inhibitor	Tumor type	Trial identifier
			Palbociclib	
Aromatase	Letrozole, anastrozole, exemestane		HR$^+$ breast cancer, HR$^+$ ovarian cancer, metastatic breast cancer, metastatic endometrial cancer	NCT04130152, NCT03054363, NCT03936270, NCT04047758, NCT02692755, NCT02806050, NCT03870919, NCT02040857, NCT04176354, NCT02028507, NCT03220178, NCT02592083, NCT02603679, NCT04256941, NCT03425838, NCT02894398, NCT02297438, NCT02730429, NCT02142868, NCT02942355
LHRH	LHRH agonists: goserelin, leuprolide		HR$^+$ breast cancer	NCT03969121, NCT03423199, NCT01723774, NCT02917005, NCT02592746, NCT03628066
ER	ER antagonists: fulvestrant, tamoxifen		HR$^+$ breast cancer, metastatic breast cancer	NCT02668666, NCT02738866, NCT03184090, NCT04526028, NCT02513394, NCT03560856, NCT02760030, NCT03079011, NCT03227328, NCT03809988, NCT02764541, NCT03007979, NCT03633331
ER	Selective estrogen receptor degraders (SERDs): G1T48, ZN-c5, SAR439859, AZD9833, GDC-9545		HR$^+$ breast cancer	NCT03455270, NCT04546009, NCT04436744, NCT04478266, NCT03560531, NCT03616587, NCT03284957, NCT03332797
ER	Selective estrogen receptor modulator (SERM): bazedoxifene		HR$^+$ breast cancer	NCT03820830, NCT02448771
Aromatase + PD-1	Letrozole, anastrozole	Pembrolizumab, nivolumab	Stage IV ER$^+$ breast cancer	NCT02778685, NCT04075604
PD-1		Nivolumab, pembrolizumab, MGA012	Liposarcoma	NCT04438824
PD-L1		Avelumab	AR$^+$ breast cancer, TNBC, ER$^+$/HER2$^-$ metastatic breast cancer	NCT04360941, NCT03147287
EGFR + PD-L1	Cetuximab	Avelumab	Squamous cell carcinoma of the head and neck	NCT03498378
HER2	Tucatinib, trastuzumab, pertuzumab, T-DM1, ZW25		HER2$^+$ breast cancer	NCT03530696, NCT03054363, NCT02448420, NCT03709082, NCT03304080, NCT02947685
EGFR/HER2	Neratinib		Advanced solid tumors with EGFR mutation/amplification, HER2 mutation/amplification, HER3/4 mutation, or KRAS mutation	NCT03065387
EGFR	Cetuximab		Metastatic colorectal cancer, squamous cell carcinoma of the head and neck	NCT03446157, NCT02499120
FGFR	Erdafitinib		ER$^+$/HER2$^-$/FGFR-amplified metastatic breast cancer	NCT03238196

continued on next page

Additional target	Inhibitor	Immune checkpoint inhibitor	Tumor type	Trial identifier
FGFR1-3	Rogaratinib		FGFR1-3$^+$/HR$^+$ breast cancer	NCT04483505
IGF-1R	Ganitumab		Ewing sarcoma	NCT04129151
VEGF1-3 receptors + PD-L1	Axitinib	Avelumab	NSCLC	NCT03386929
RAF	Sorafenib		Leukemia	NCT03132454
MEK	PD-0325901, binimetinib		KRAS mutant NSCLC, TNBC, KRAS and NRAS mutant metastatic or unresectable colorectal cancer	NCT02022982, NCT03170206, NCT04494958, NCT03981614
ERK	Ulixertinib		Advanced pancreatic cancer and other solid tumors	NCT03454035
PI3K	Copanlisib		HR$^+$ breast cancer	NCT03128619
PI3K	Taselisib, pictilisib, GDC-0077		PIK3CA mutant advanced solid tumors, PIK3CA mutant and HR$^+$ breast cancer	NCT02389842, NCT04191499, NCT03006172
PI3K/mTOR	Gedatolisib		Metastatic breast cancer, advanced squamous cell lung, pancreatic, head and neck cancer and other solid tumors	NCT02684032, NCT03065062, NCT02626507
mTOR	Everolimus, vistusertib		HR$^+$ breast cancer	NCT02871791
AKT	Ipatasertib		HR$^+$ breast cancer, metastatic breast cancer, metastatic gastrointestinal tumors, NSCLC	NCT03959891, NCT04060862, NCT04591431
BTK	Ibrutinib		Mantle cell lymphoma	NCT03478514
BCL-2	Venetoclax		ER$^+$/BCL-2$^+$ advanced or metastatic breast cancer	NCT03900884
AR	AR antagonists: bicalutamide		AR$^+$ metastatic breast cancer	NCT02605486
Lysosome + aromatase	Hydroxychloroquine + letrozole		ER$^+$ breast cancer	NCT03774472
Proliferating cells	Standard chemotherapy		Stage IV ER$^+$ breast cancer	NCT03355157
Proliferating cells	Radiation		Stage IV ER$^+$ breast cancer	NCT03870919, NCT03691493, NCT04605562
BCR-ABL	Bosutinib		HR$^+$ breast cancer	NCT03854903
Ribociclib				
Aromatase	Letrozole, anastrozole, exemestane		HR$^+$ breast cancer, metastatic breast cancer, ovarian cancer	NCT04256941, NCT03425838, NCT03822468, NCT02712723, NCT03673124, NCT02941926, NCT03248427, NCT03671330, NCT02333370, NCT01958021, NCT03425838
LHRH	LHRH agonists: goserelin, leuprolide		HR$^+$ breast cancer	NCT03944434
ER	ER antagonists: fulvestrant		HR$^+$ breast cancer, advanced breast cancer	NCT03227328, NCT02632045, NCT02632045, NCT03555877
PD-1		Spartalizumab	Breast cancer and ovarian cancer, recurrent and/or metastatic head and neck squamous cell carcinoma, melanoma	NCT03294694, NCT04213404, NCT03484923
HER2	Trastuzumab, pertuzumab, T-DM1		HER2$^+$ breast cancer	NCT03913234, NCT02657343
EGFR	Nazartinib (EGF816)		EGFR mutant NSCLC	NCT03333343
RAF	Encorafenib, LXH254		NSCLC, BRAF mutant melanoma	NCT02974725, NCT03333343, NCT04417621, NCT02159066

continued on next page

Additional target	Inhibitor	Immune checkpoint inhibitor	Tumor type	Trial identifier
MEK	Binimetinib		BRAF V600-dependent advanced solid tumors, melanoma	NCT01543698, NCT02159066
PI3K	Alpelisib		Breast cancer with PIK3CA mutation	NCT03439046
mTOR	Everolimus		Advanced dedifferentiated liposarcoma, leiomyosarcoma, glioma, astrocytoma, glioblastoma, endometrial carcinoma, pancreatic cancer, neuroendocrine tumors	NCT03114527, NCT03355794, NCT03834740, NCT03008408, NCT02985125, NCT03070301
mTOR + inflammation	Everolimus + dexamethasone		ALL	NCT03740334
SHP2	TNO155		Advanced solid tumors	NCT04000529
AR	AR antagonists: bicalutamide, enzalutamide		TNBC, metastatic prostate carcinoma	NCT03090165, NCT02555189
HDAC	Belinostat		TNBC, ovarian cancer	NCT04315233
proliferating cells	Standard chemotherapy		Ovarian cancer, metastatic solid tumors, soft tissue sarcoma, hepatocellular carcinoma	NCT03056833, NCT03237390, NCT03009201, NCT02524119
Abemaciclib				
Aromatase	Letrozole, anastrozole, exemestane		HR$^+$ breast cancer, metastatic breast cancer, endometrial cancer	NCT04256941, NCT03425838, NCT04227327, NCT04393285, NCT04305236, NCT03643510, NCT03675893, NCT04352777, NCT04293393, NCT02057133
ER	ER antagonists: fulvestrant		Advanced breast cancer, low-grade serous ovarian cancer	NCT03227328, NCT03531645, NCT04158362, NCT01394016
PD-1		Nivolumab, pembrolizumab	Head and neck cancer, gastroesophageal cancer, NSCLC, HR$^+$ breast cancer	NCT04169074, NCT03655444, NCT03997448, NCT02779751
ER + PD-L1	ER antagonists: fulvestrant	Atezolizumab	HR$^+$ breast cancer, metastatic breast cancer	NCT03280563
AKT + ER + PD-L1	Ipatasertib + ER antagonists: fulvestrant	Atezolizumab	HR$^+$ breast cancer	NCT03280563
PD-L1		LY3300054	Advanced solid tumors	NCT02791334
HER2	Trastuzumab		HER2$^+$ metastatic breast cancer	NCT04351230
Receptor tyrosine kinases	Sunitinib		Metastatic renal cell carcinoma	NCT03905889
IGF-1/IGF-2	Xentuzumab		HR$^+$ breast cancer	NCT03099174
VEGF-A	Bevacizumab		Glioblastoma	NCT04074785
PI3K	Copanlisib		HR$^+$ breast cancer, metastatic breast cancer	NCT03939897
PI3K/mTOR	LY3023414		Metastatic cancer	NCT01655225
ERK1/2	LY3214996		tumors with ERK1/2 mutations, glioblastoma, metastatic cancer	NCT04534283, NCT04391595, NCT02857270
Trilaciclib				
Proliferating cells	Chemotherapy		SCLC: This trial evaluates the potential clinical benefit of trilaciclib in preventing chemotherapy-induced myelosuppression in patients receiving chemotherapy	NCT04504513

continued on next page

Additional target	Inhibitor	Immune checkpoint inhibitor	Tumor type	Trial identifier
Proliferating cells + PD-L1	Carboplatin + etoposide	Atezolizumab	SCLC: This trial investigates the potential clinical benefit of trilaciclib in preserving the bone marrow and the immune system, and enhancing antitumor efficacy when administered with chemotherapy	NCT03041311
Proliferating cells	Topotecan		SCLC: This trial investigates the potential clinical benefit of trilaciclib in preserving the bone marrow and the immune system, and enhancing the antitumor efficacy of chemotherapy when administered prior to chemotherapy	NCT02514447
Proliferating cells	Carboplatin + gemcitabine		Metastatic TNBC: This study investigates the potential clinical benefit of trilaciclib in preserving the bone marrow and the immune system, and enhancing the antitumor efficacy of chemotherapy when administered prior to chemotherapy	NCT02978716
Lerociclib				
ER	ER antagonist: fulvestrant		HR$^+$/HER2$^-$ metastatic breast cancer	NCT02983071
EGFR	Osimertinib		EGFR mutant NSCLC	NCT03455829
SHR6390				
ER	ER antagonist: fulvestrant		HR$^+$/HER2$^-$ recurrent/ metastatic breast cancer	NCT03481998
Aromatase	Letrozole, anastrozole		HR$^+$/HER2$^-$ recurrent/ metastatic breast cancer	NCT03966898, NCT03772353
EGFR/HER2	Pyrotinib		HER2$^+$ gastric cancer, HER2$^+$ metastatic breast cancer	NCT04095390, NCT03993964
AR	AR antagonists: SHR3680		metastatic TNBC	NCT03805399
PF-06873600				
Endocrine therapy	Single agent and then in combination with endocrine therapy		HR$^+$/HER2$^-$ metastatic breast cancer, ovarian and fallopian tube cancer, TNBC and other tumors	NCT03519178
FCN-473c				
Aromatase	Letrozole		ER$^+$/HER2$^-$ advanced breast cancer	NCT04488107

combination with CDK4/6 inhibitors blocked proliferation of CDK4/6 inhibitor–resistant cancer cells (*111*, *141*, *158–161*). Recently, two CDK2-specific inhibitors, PF-07104091 (*163*) and BLU0298 (*164*), have been reported. PF-07104091 is now being tested in a phase 2 clinical trial in combination with palbociclib plus antiestrogens. Another recent study identified a novel compound, PF-3600, that inhibits CDK4/6 and CDK2 (*165*). PF3600 had potent antitumor effects against xenograft models of intrinsic and acquired resistance to CDK4/6 inhibition

(*165*). A phase 2 clinical trial is currently evaluating this compound as a single agent and in combination with endocrine therapy in patients with HR$^+$/HER2$^-$ breast cancer and other cancer types.

Whole-exome sequencing of 59 HR$^+$/HER2$^-$ metastatic breast tumors from patients treated with CDK4/6 inhibitors and anti-estrogens revealed eight alterations that likely conferred resistance: *RB1* loss; amplification of *CCNE2* or *AURKA*; activating mutations or amplification of *AKT1*, *FGFR2*, or *ERBB2*; activating muta-

tions in *RAS* genes; and loss of ER expression. The frequent activation of AURKA (in 27% of resistant tumors) raises the possibility of combining CDK4/6 inhibitors with inhibitors of Aurora A kinase to overcome resistance (*145*).

In contrast to ER$^+$ mammary carcinomas, TNBCs are overall resistant to CDK4/6 inhibition (*45*). A subset of TNBCs display high numbers of lysosomes, which causes sequestration of CDK4/6 inhibitors into the expanded lysosomal compartment, thereby preventing their action on nuclear CDK4/6. Preclinical studies
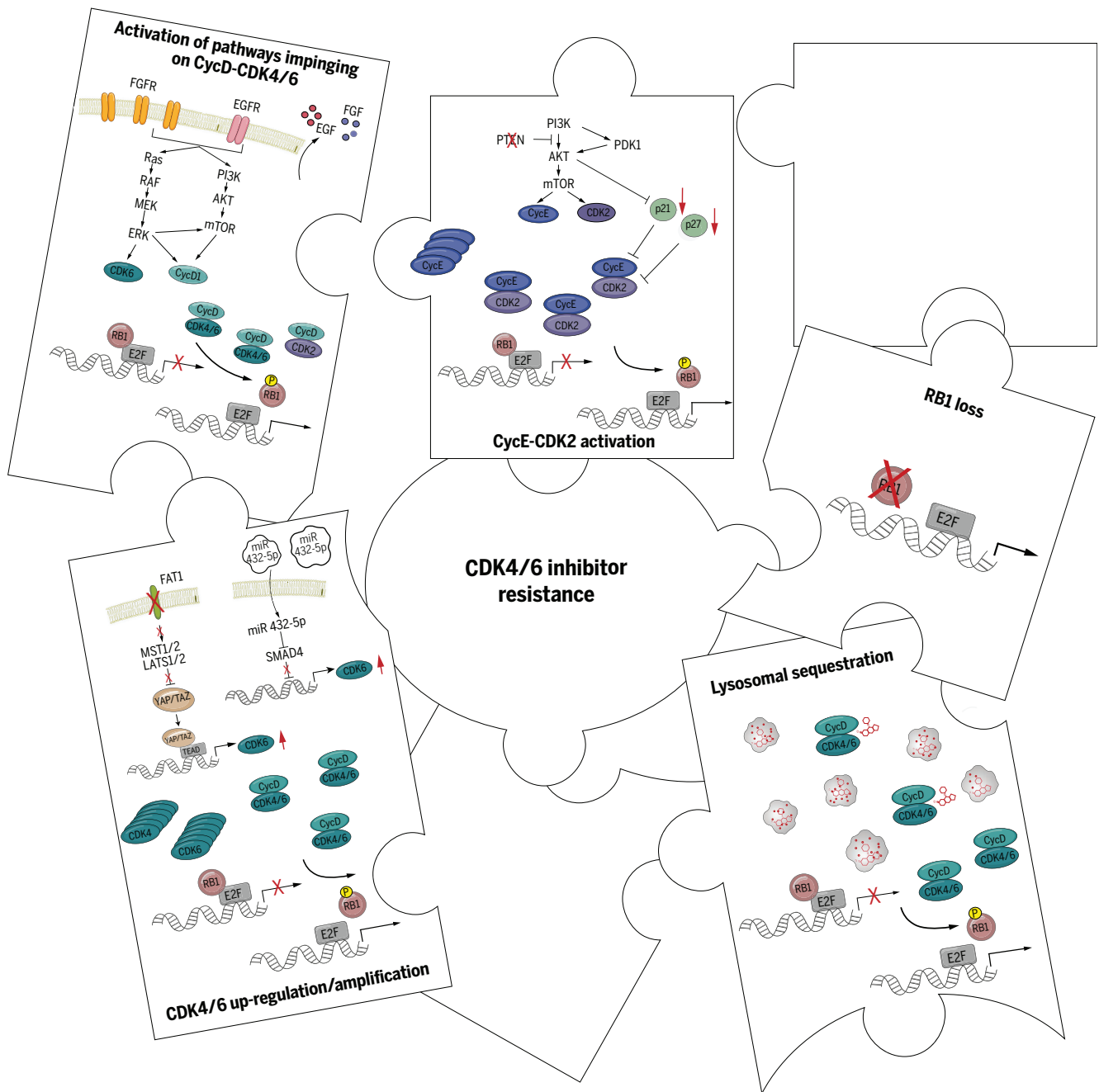
Fig. 3. Mechanisms of cancer cell resistance to CDK4/6 inhibition. Known mechanisms include loss of RB1, activation of pathways impinging on CycD-CDK4/6, amplification of the *CDK4/6* genes and overexpression of CDK6 protein, activation of CycE-CDK2, and lysosomal sequestration of CDK4/6 inhibitors. Blank pieces of the puzzle denote additional mechanisms that remain to be discovered.

revealed that lysosomotropic agents that reverse the lysosomal sequestration (such as chloroquine, azithromycin, or siramesine) render TNBC cells fully sensitive to CDK4/6 inhibition (*71*, *111*). These observations now need to be tested in clinical trials for TNBC patients.

Outlook

Although D-cyclins and CDK4/6 were discovered 30 years ago, several aspects of cyclin D–CDK4/6 biology, such as their role in

antitumor immunity, are only now starting to be appreciated. The full range of cyclin D–CDK4/6 functions in tumor cells remains unknown. It is likely that these kinases play a much broader role in cancer cells than is currently appreciated. Hence, the impact of CDK4/6 inhibition on various aspects of tumorigenesis requires further study. Also, treatment of patients with CDK4/6 inhibitors likely affects several aspects of host physiology, which may be relevant to cancer progression.

In the next years, we will undoubtedly witness the development and testing of new CDK4/6 inhibitors. Because activation of CDK2 represents a frequent CDK4/6 inhibitor resistance mechanism, compounds that inhibit CDK4/6 and CDK2 may prevent or delay the development of resistance. Conversely, selective compounds that inhibit CDK4 but not CDK6 may allow more aggressive dosing, as they are expected not to result in bone marrow toxicity caused by CDK6 inhibition. New, less basic CDK4/6 inhibitor compounds (*111*) may

escape lysosomal sequestration and may be efficacious against resistant cancer types such as TNBC. Degrader compounds, which induce proteolysis of cyclin D rather than inhibit cyclin D–CDK4/6 kinase, may have superior properties, as they would extinguish both CDK4/6-dependent and -independent functions of D-cyclins in tumorigenesis. Moreover, dissolution of cyclin D–CDK4/6 complexes is expected to liberate KIP/CIP inhibitors, which would then inhibit CDK2. D-cyclins likely play CDK-independent functions in tumorigenesis—for example, by regulating gene expression (166). However, their role in tumor biology and the utility of targeting these functions for cancer treatment remain largely unexplored.

An important challenge will be to test and identify combinatorial treatments involving CDK4/6 inhibitors for the treatment of different tumor types. CDK4/6 inhibitors trigger cell cycle arrest of tumor cells and, in some cases, senescence. It will be essential to identify combination treatments that convert CDK4/6 inhibitors from cytostatic compounds to cytotoxic ones, which would unleash the killing of tumor cells. Genome-wide high-throughput screens along with analyses of mouse cancer models and PDXs will help to address this point. Another largely unexplored area of cyclin D–CDK4/6 biology is the possible involvement of these proteins in other pathologies, such as metabolic disorders. Research in this area may extend the use of CDK4/6 inhibitors to treatment of other diseases. All these unresolved questions ensure that CDK4/6 biology will remain an active area of basic, translational, and clinical research for several years to come.

REFERENCES AND NOTES

1. H. Matsushime, M. F. Roussel, R. A. Ashmun, C. J. Sherr, Colony-stimulating factor 1 regulates novel cyclins during the G1 phase of the cell cycle. *Cell* 65, 701–713 (1991). doi: 10.1016/0092-8674(91)90101-4; pmid: 1827757

2. Y. Xiong, T. Connolly, B. Futcher, D. Beach, Human D-type cyclin. *Cell* 65, 691–699 (1991). doi: 10.1016/0092-8674(91)90100-D; pmid: 1827756

3. T. Motokura et al., A novel cyclin encoded by a bcl1-linked candidate oncogene. *Nature* 350, 512–515 (1991). doi: 10.1038/350512a0; pmid: 1826542

4. D. A. Withers et al., Characterization of a candidate bcl-1 gene. *Mol. Cell. Biol.* 11, 4846–4853 (1991). pmid: 1833629

5. M. Malumbres, M. Barbacid, To cycle or not to cycle: A critical decision in cancer. *Nat. Rev. Cancer* 1, 222–231 (2001). doi: 10.1038/35106065; pmid: 11902577

6. C. J. Sherr, J. M. Roberts, CDK inhibitors: Positive and negative regulators of G1-phase progression. *Genes Dev.* 13, 1501–1512 (1999). doi: 10.1101/gad.13.12.1501; pmid: 10385618

7. M. Malumbres et al., Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell* 118, 493–504 (2004). doi: 10.1016/j.cell.2004.08.002; pmid: 15315761

8. K. Kozar et al., Mouse development and cell proliferation in the absence of D-cyclins. *Cell* 118, 477–491 (2004). doi: 10.1016/j.cell.2004.07.025; pmid: 15315760

9. T. Otto, P. Sicinski, Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer* 17, 93–115 (2017). doi: 10.1038/nrc.2016.138; pmid: 28127048

10. M. Cheng et al., The p21(Cip1) and p27(Kip1) CDK 'inhibitors' are essential activators of cyclin D-dependent kinases in murine fibroblasts. *EMBO J.* 18, 1571–1583 (1999). doi: 10.1093/emboj/18.6.1571; pmid: 10075928

11. M. K. James, A. Ray, D. Leznova, S. W. Blain, Differential modification of p27Kip1 controls its cyclin D-cdk4 inhibitory activity. *Mol. Cell. Biol.* 28, 498–510 (2008). doi: 10.1128/MCB.02171-06; pmid: 17908796

12. A. Ray, M. K. James, S. Larochelle, R. P. Fisher, S. W. Blain, p27Kip1 inhibits cyclin D-cyclin-dependent kinase 4 by two independent modes. *Mol. Cell. Biol.* 29, 986–999 (2009). doi: 10.1128/MCB.00898-08; pmid: 19075005

13. P. Patel et al., Brk/Protein tyrosine kinase 6 phosphorylates p27KIP1, regulating the activity of cyclin D-cyclin-dependent kinase 4. *Mol. Cell. Biol.* 35, 1506–1522 (2015). doi: 10.1128/MCB.01206-14; pmid: 25733683

14. S. W. Blain, Switching cyclin D-Cdk4 kinase activity on and off. *Cell Cycle* 7, 892–898 (2008). doi: 10.4161/cc.7.7.5637; pmid: 18414028

15. J. A. Diehl, M. Cheng, M. F. Roussel, C. J. Sherr, Glycogen synthase kinase-3β regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev.* 12, 3499–3511 (1998). doi: 10.1101/gad.12.22.3499; pmid: 9832503

16. J. R. Alt, J. L. Cleveland, M. Hannink, J. A. Diehl, Phosphorylation-dependent regulation of cyclin D1 nuclear export and cyclin D1-dependent cellular transformation. *Genes Dev.* 14, 3102–3114 (2000). doi: 10.1101/gad.854900; pmid: 11124803

17. S. Qie, J. A. Diehl, Cyclin D1, cancer progression, and opportunities in cancer treatment. *J. Mol. Med.* 94, 1313–1326 (2016). doi: 10.1007/s00109-016-1475-3; pmid: 27695879

18. A. Yoshida et al., Fbxl8 suppresses lymphoma growth and hematopoietic transformation through degradation of cyclin D3. *Oncogene* 40, 292–306 (2021). doi: 10.1038/s41388-020-01532-4; pmid: 33122824

19. T. Kanie et al., Genetic reevaluation of the role of F-box proteins in cyclin D1 degradation. *Mol. Cell. Biol.* 32, 590–605 (2012). doi: 10.1128/MCB.06570-11; pmid: 22124152

20. A. C. Chaikovsky et al., The AMBRA1 E3 ligase adaptor regulates the stability of cyclin D. *Nature* 592, 794–798 (2021). doi: 10.1038/s41586-021-03474-7; pmid: 33854239

21. E. Maiani et al., AMBRA1 regulates cyclin D to guard S-phase entry and genomic integrity. *Nature* 592, 799–803 (2021). doi: 10.1038/s41586-021-03422-5; pmid: 33854232

22. D. Simoneschi et al., CRL4^AMBRA1 is a master regulator of D-type cyclins. *Nature* 592, 789–793 (2021). doi: 10.1038/s41586-021-03445-y; pmid: 33854235

23. M. Hall, G. Peters, Genetic alterations of cyclins, cyclin-dependent kinases, and Cdk inhibitors in human cancer. *Adv. Cancer Res.* 68, 67–108 (1996). doi: 10.1016/S0065-230X(08)60352-8; pmid: 8712071

24. A. Arnold, A. Papanikolaou, Cyclin D1 in breast cancer pathogenesis. *J. Clin. Oncol.* 23, 4215–4224 (2005). doi: 10.1200/JCO.2005.05.064; pmid: 15961768

25. T. C. Wang et al., Mammary hyperplasia and carcinoma in MMTV-cyclin D1 transgenic mice. *Nature* 369, 669–671 (1994). doi: 10.1038/369669a0; pmid: 8208295

26. M. W. Landis, B. S. Pawlyk, T. Li, P. Sicinski, P. W. Hinds, Cyclin D1-dependent kinase activity in murine development and mammary tumorigenesis. *Cancer Cell* 9, 13–22 (2006). doi: 10.1016/j.ccr.2005.12.019; pmid: 16413468

27. H. K. Reddy et al., Cyclin-dependent kinase 4 expression is essential for neu-induced breast tumorigenesis. *Cancer Res.* 65, 10174–10178 (2005). doi: 10.1158/0008-5472.CAN-05-2639; pmid: 16288002

28. Q. Yu, Y. Geng, P. Sicinski, Specific protection against breast cancers by cyclin D1 ablation. *Nature* 411, 1017–1021 (2001). doi: 10.1038/35082500; pmid: 11429595

29. Q. Yu et al., Requirement for CDK4 kinase function in breast cancer. *Cancer Cell* 9, 23–32 (2006). doi: 10.1016/j.ccr.2005.12.012; pmid: 16413469

30. Y. J. Choi et al., The requirement for cyclin D function in tumor maintenance. *Cancer Cell* 22, 438–451 (2012). doi: 10.1016/j.ccr.2012.09.015; pmid: 23079655

31. M. Puyol et al., A synthetic lethal interaction between K-Ras oncogenes and Cdk4 unveils a therapeutic strategy for non-small cell lung carcinoma. *Cancer Cell* 18, 63–73 (2010). doi: 10.1016/j.ccr.2010.05.025; pmid: 20609353

32. A. M. Narasimha et al., Cyclin D activates the Rb tumor suppressor by mono-phosphorylation. *eLife* 3, e02872 (2014). doi: 10.7554/eLife.02872; pmid: 24876129

33. M. Chung et al., Transient Hysteresis in CDK4/6 Activity Underlies Passage of the Restriction Point in G1. *Mol. Cell* 76, 562–573.e4 (2019). doi: 10.1016/j.molcel.2019.08.020; pmid: 31543423

34. B. R. Topacio et al., Cyclin D-Cdk4,6 Drives Cell-Cycle Progression via the Retinoblastoma Protein's C-Terminal Helix. *Mol. Cell* 74, 758–770.e4 (2019). doi: 10.1016/j.molcel.2019.03.020; pmid: 30982746

35. L. Anders et al., A systematic screen for CDK4/6 substrates links FOXM1 phosphorylation to senescence suppression in cancer cells. *Cancer Cell* 20, 620–634 (2011). doi: 10.1016/j.ccr.2011.10.001; pmid: 22094256

36. I. Matsuura et al., Cyclin-dependent kinases regulate the antiproliferative function of Smads. *Nature* 430, 226–231 (2004). doi: 10.1038/nature02650; pmid: 15241418

37. I. The et al., Rb and FZR1/Cdh1 determine CDK4/6-cyclin D requirement in C. elegans and human cancer cells. *Nat. Commun.* 6, 5906 (2015). doi: 10.1038/ncomms6906; pmid: 25562820

38. J. Romero-Pozuelo, G. Figlia, O. Kaya, A. Martin-Villalba, A. A. Teleman, Cdk4 and Cdk6 Couple the Cell-Cycle Machinery to Cell Growth via mTORC1. *Cell Rep.* 31, 107504 (2020). doi: 10.1016/j.celrep.2020.03.068; pmid: 32294430

39. P. Aggarwal et al., Nuclear cyclin D1/CDK4 kinase regulates CUL4 expression and triggers neoplastic growth via activation of the PRMT5 methyltransferase. *Cancer Cell* 18, 329–340 (2010). doi: 10.1016/j.ccr.2010.08.012; pmid: 20951943

40. Y. Li et al., PRMT5 is required for lymphomagenesis triggered by multiple oncogenic drivers. *Cancer Discov.* 5, 288–303 (2015). doi: 10.1158/2159-8290.CD-14-0625; pmid: 25582697

41. S. AbuHammad et al., Regulation of PRMT5-MDM4 axis is critical in the response to CDK4/6 inhibitors in melanoma. *Proc. Natl. Acad. Sci. U.S.A.* 116, 17990–18000 (2019). doi: 10.1073/pnas.1901323116; pmid: 31439820

42. F. Bellutti et al., CDK6 Antagonizes p53-Induced Responses during Tumorigenesis. *Cancer Discov.* 8, 884–897 (2018). doi: 10.1158/2159-8290.CD-17-0912; pmid: 29899063

43. I. Z. Uras et al., Palbociclib treatment of FLT3-ITD+ AML cells uncovers a kinase-dependent transcriptional regulation of FLT3 and PIM1 by CDK6. *Blood* 127, 2890–2902 (2016). doi: 10.1182/blood-2015-11-683581; pmid: 27099147

44. K. Z. Guiley et al., p27 allosterically activates cyclin-dependent kinase 4 and antagonizes palbociclib inhibition. *Science* 366, eaaw2106 (2019). doi: 10.1126/science.aaw2106; pmid: 31831640

45. R. S. Finn et al., PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro. *Breast Cancer Res.* 11, R77 (2009). doi: 10.1186/bcr2419; pmid: 19874578

46. D. W. Fry et al., Specific inhibition of cyclin-dependent kinase 4/6 by PD 0332991 and associated antitumor activity in human tumor xenografts. *Mol. Cancer Ther.* 3, 1427–1438 (2004). pmid: 15542782

47. N. S. Zainal et al., Effects of palbociclib on oral squamous cell carcinoma and the role of PIK3CA in conferring resistance. *Cancer Biol. Med.* 16, 264–275 (2019). doi: 10.20892/j.issn.2095-3941.2018.0257; pmid: 31516747

48. R. Saab et al., Pharmacologic inhibition of cyclin-dependent kinase 4/6 activity arrests proliferation in myoblasts and rhabdomyosarcoma-derived cells. *Mol. Cancer Ther.* 5, 1299–1308 (2006). doi: 10.1158/1535-7163.MCT-05-0383; pmid: 16731763

49. S. Paternot, B. Colleoni, X. Bisteau, P. P. Roger, The CDK4/CDK6 inhibitor PD0332991 paradoxically stabilizes activated cyclin D3-CDK4/6 complexes. *Cell Cycle* 13, 2879–2888 (2014). doi: 10.4161/15384101.2014.946841; pmid: 25486476

50. Y. Geng et al., Kinase-independent function of E-type cyclins in liver cancer. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1015–1020 (2018). doi: 10.1073/pnas.1711477115; pmid: 29339491

51. O. Tetsu, F. McCormick, Proliferation of cancer cells despite CDK2 inhibition. *Cancer Cell* 3, 233–245 (2003). doi: 10.1016/S1535-6108(03)00053-9; pmid: 12676582

52. S. T. Hallett et al., Differential Regulation of G1 CDK Complexes by the Hsp90-Cdc37 Chaperone System. *Cell Rep.* 21, 1386–1398 (2017). doi: 10.1016/j.celrep.2017.10.042; pmid: 29091774

53. S. Polier et al., ATP-competitive inhibitors block protein kinase recruitment to the Hsp90-Cdc37 system. *Nat. Chem. Biol.* 9, 307–312 (2013). doi: 10.1038/nchembio.1212; pmid: 23502424

54. J. L. Dean, C. Thangavel, A. K. McClendon, C. A. Reed, E. S. Knudsen, Therapeutic CDK4/6 inhibition in breast

cancer: Key mechanisms of response and failure. *Oncogene* **29**, 4018–4032 (2010). doi: 10.1038/onc.2010.154; pmid: 20473330

55. J. Lukas, J. Bartkova, M. Rohde, M. Strauss, J. Bartek, Cyclin D1 is dispensable for G1 control in retinoblastoma gene-deficient cells independently of cdk4 activity. *Mol. Cell. Biol.* **15**, 2600–2611 (1995). doi: 10.1128/MCB.15.5.2600; pmid: 7739541

56. J. Lukas *et al.*, Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature* **375**, 503–506 (1995). doi: 10.1038/375503a0; pmid: 7777060

57. R. H. Medema, R. E. Herrera, F. Lam, R. A. Weinberg, Growth suppression by p16ink4 requires functional retinoblastoma protein. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6289–6293 (1995). doi: 10.1073/pnas.92.14.6289; pmid: 7603984

58. X. Gong *et al.*, Genomic Aberrations that Activate D-type Cyclins Are Associated with Enhanced Sensitivity to the CDK4 and CDK6 Inhibitor Abemaciclib. *Cancer Cell* **32**, 761–776.e6 (2017). doi: 10.1016/j.ccell.2017.11.006; pmid: 29232554

59. S. Kim *et al.*, The potent and selective cyclin-dependent kinases 4 and 6 inhibitor ribociclib (LEE011) is a versatile combination partner in preclinical cancer models. *Oncotarget* **9**, 35226–35240 (2018). doi: 10.18632/oncotarget.26215; pmid: 30443290

60. J. E. Bisi, J. A. Sorrentino, P. J. Roberts, F. X. Tavares, J. C. Strum, Preclinical Characterization of G1T28: A Novel CDK4/6 Inhibitor for Reduction of Chemotherapy-Induced Myelosuppression. *Mol. Cancer Ther.* **15**, 783–793 (2016). doi: 10.1158/1535-7163.MCT-15-0775; pmid: 26826116

61. J. Rader *et al.*, Dual CDK4/CDK6 inhibition induces cell-cycle arrest and senescence in neuroblastoma. *Clin. Cancer Res.* **19**, 6173–6182 (2013). doi: 10.1158/1078-0432.CCR-13-1675; pmid: 24045179

62. R. Torres-Guzmán *et al.*, Preclinical characterization of abemaciclib in hormone receptor positive breast cancer. *Oncotarget* **8**, 69493–69507 (2017). doi: 10.18632/oncotarget.17778; pmid: 29050219

63. K. Michaud *et al.*, Pharmacologic inhibition of cyclin-dependent kinases 4 and 6 arrests the growth of glioblastoma multiforme intracranial xenografts. *Cancer Res.* **70**, 3228–3238 (2010). doi: 10.1158/0008-5472.CAN-09-4559; pmid: 20354191

64. T. J. Raub *et al.*, Brain Exposure of Two Selective Dual CDK4 and CDK6 Inhibitors and the Antitumor Activity of CDK4 and CDK6 Inhibition in Combination with Temozolomide in an Intracranial Glioblastoma Xenograft. *Drug Metab. Dispos.* **43**, 1360–1371 (2015). doi: 10.1124/dmd.114.062745; pmid: 26149830

65. F. Long *et al.*, Preclinical characterization of SHR6390, a novel CDK 4/6 inhibitor, in vitro and in human tumor xenograft models. *Cancer Sci.* **110**, 1420–1430 (2019). doi: 10.1111/cas.13957; pmid: 30724426

66. S. Lin *et al.*, FCN-437: A novel, potent and selective oral inhibitor of CDK4/6 for the treatment of solid tumors. *Cancer Res.* **79** (suppl. 13), 4425 (2019). doi: 10.1158/1538-7445.AM2019-4425

67. L. Yin *et al.*, A highly potent CDK4/6 inhibitor was rationally designed to overcome blood brain barrier in glioblastoma therapy. *Eur. J. Med. Chem.* **144**, 1–28 (2018). doi: 10.1016/j.ejmech.2017.12.003; pmid: 29247857

68. C. Rubio *et al.*, CDK4/6 Inhibitor as a Novel Therapeutic Approach for Advanced Bladder Cancer Independently of RB1 Status. *Clin. Cancer Res.* **25**, 390–402 (2019). doi: 10.1158/1078-0432.CCR-18-0685; pmid: 30242024

69. D. L. Burkhart, J. Sage, Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nat. Rev. Cancer* **8**, 671–682 (2008). doi: 10.1038/nrc2399; pmid: 18650841

70. A. Smirnov *et al.*, FOXM1 regulates proliferation, senescence and oxidative stress in keratinocytes and cancer cells. *Aging* **8**, 1384–1397 (2016). doi: 10.18632/aging.100988; pmid: 27385468

71. S. Vijayaraghavan *et al.*, CDK4/6 and autophagy inhibitors synergistically induce senescence in Rb positive cytoplasmic cyclin E negative cancers. *Nat. Commun.* **8**, 15916 (2017). doi: 10.1038/ncomms15916; pmid: 28653662

72. A. Yoshida, E. K. Lee, J. A. Diehl, Induction of Therapeutic Senescence in Vemurafenib-Resistant Melanoma by Extended Inhibition of CDK4/6. *Cancer Res.* **76**, 2990–3002 (2016). doi: 10.1158/0008-5472.CAN-15-2931; pmid: 26988987

73. M. Kovatcheva *et al.*, MDM2 turnover and expression of ATRX determine the choice between quiescence and senescence in

response to CDK4 inhibition. *Oncotarget* **6**, 8226–8243 (2015). doi: 10.18632/oncotarget.3364; pmid: 25803170

74. M. E. Klein *et al.*, PDLIM7 and CDH18 regulate the turnover of MDM2 during CDK4/6 inhibitor therapy-induced senescence. *Oncogene* **37**, 5066–5078 (2018). doi: 10.1038/s41388-018-0332-y; pmid: 29789718

75. W. R. Wiedemeyer *et al.*, Pattern of retinoblastoma pathway inactivation dictates response to CDK4/6 inhibition in GBM. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11501–11506 (2010). doi: 10.1073/pnas.1001613107; pmid: 20534551

76. S. R. S. Gottesman *et al.*, Tyrosine Phosphorylation of p27Kip1 Correlates with Palbociclib Responsiveness in Breast Cancer Tumor Cells Grown in Explant Culture. *Mol. Cancer Res.* **17**, 669–675 (2019). doi: 10.1158/1541-7786.MCR-18-0188; pmid: 30559257

77. E. Raspé *et al.*, CDK4 phosphorylation status and a linked gene expression profile predict sensitivity to palbociclib. *EMBO Mol. Med.* **9**, 1052–1066 (2017). doi: 10.15252/emmm.201607084; pmid: 28566333

78. R. S. Finn *et al.*, The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): A randomised phase 2 study. *Lancet Oncol.* **16**, 25–35 (2015). doi: 10.1016/S1470-2045(14)71159-3; pmid: 25524798

79. R. S. Finn *et al.*, Biomarker Analyses of Response to Cyclin-Dependent Kinase 4/6 Inhibition and Endocrine Therapy in Women with Treatment-Naïve Metastatic Breast Cancer. *Clin. Cancer Res.* **26**, 110–121 (2020). doi: 10.1158/1078-0432.CCR-19-0751; pmid: 31527167

80. N. C. Turner *et al.*, Cyclin E1 Expression and Palbociclib Efficacy in Previously Treated Hormone Receptor-Positive Metastatic Breast Cancer. *J. Clin. Oncol.* **37**, 1169–1178 (2019). doi: 10.1200/JCO.18.00925; pmid: 30807234

81. Z. Li *et al.*, Loss of the FAT1 Tumor Suppressor Promotes Resistance to CDK4/6 Inhibitors via the Hippo Pathway. *Cancer Cell* **34**, 893–905.e8 (2018). doi: 10.1016/j.ccell.2018.11.006; pmid: 30537512

82. M. Cheng, V. Sexl, C. J. Sherr, M. F. Roussel, Assembly of cyclin D-dependent kinase and titration of p27Kip1 regulated by mitogen-activated protein kinase (MEK). *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1091–1096 (1998). doi: 10.1073/pnas.95.3.1091; pmid: 9448290

83. E. A. Klein, R. K. Assoian, Transcriptional regulation of the cyclin D1 gene at a glance. *J. Cell Sci.* **121**, 3853–3857 (2008). doi: 10.1242/jcs.039131; pmid: 19020303

84. J. N. Lavoie, G. L'Allemain, A. Brunet, R. Müller, J. Pouysségur, Cyclin D1 expression is regulated positively by the p42/p44MAPK and negatively by the p38/HOGMAPK pathway. *J. Biol. Chem.* **271**, 20608–20616 (1996). doi: 10.1074/jbc.271.34.20608; pmid: 8702807

85. J. Lukas, J. Bartkova, J. Bartek, Convergence of mitogenic signalling cascades from diverse classes of receptors at the cyclin D-cyclin-dependent kinase-pRb-controlled G1 checkpoint. *Mol. Cell. Biol.* **16**, 6917–6925 (1996). doi: 10.1128/MCB.16.12.6917; pmid: 8943347

86. R. C. Muise-Helmericks *et al.*, Cyclin D expression is controlled post-transcriptionally via a phosphatidylinositol 3-kinase/Akt-dependent pathway. *J. Biol. Chem.* **273**, 29864–29872 (1998). doi: 10.1074/jbc.273.45.29864; pmid: 9792703

87. M. Pek *et al.*, Oncogenic KRAS-associated gene signature defines co-targeting of CDK4/6 and MEK as a viable therapeutic strategy in colorectal cancer. *Oncogene* **36**, 4975–4986 (2017). doi: 10.1038/onc.2017.120; pmid: 28459468

88. S. R. Vora *et al.*, CDK 4/6 inhibitors sensitize PIK3CA mutant breast cancer to PI3K inhibitors. *Cancer Cell* **26**, 136–149 (2014). doi: 10.1016/j.ccr.2014.05.020; pmid: 25002028

89. C. A. Martin *et al.*, Palbociclib synergizes with BRAF and MEK inhibitors in treatment naïve melanoma but not after the development of BRAF inhibitor resistance. *Int. J. Cancer* **142**, 2139–2152 (2018). doi: 10.1002/ijc.31220; pmid: 29243224

90. J. L. F. Teh *et al.*, *In Vivo* E2F Reporting Reveals Efficacious Schedules of MEK1/2-CDK4/6 Targeting and mTOR-S6 Resistance Mechanisms. *Cancer Discov.* **8**, 568–581 (2018). doi: 10.1158/2159-8290.CD-17-0699; pmid: 29496664

91. A. Dall'Acqua *et al.*, CDK6 protects epithelial ovarian cancer from platinum-induced death via FOXO3 regulation. *EMBO Mol. Med.* **9**, 1415–1433 (2017). doi: 10.15252/emmm.201607012; pmid: 28778953

92. H. Gao *et al.*, High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* **21**, 1318–1325 (2015). doi: 10.1038/nm.3954; pmid: 26479923

93. V. Yadav *et al.*, The CDK4/6 inhibitor LY2835219 overcomes vemurafenib resistance resulting from MAPK reactivation and cyclin D1 upregulation. *Mol. Cancer Ther.* **13**, 2253–2263 (2014). doi: 10.1158/1535-7163.MCT-14-0257; pmid: 25122067

94. A. C. Wood *et al.*, Dual *ALK* and *CDK4/6* Inhibition Demonstrates Synergy against Neuroblastoma. *Clin. Cancer Res.* **23**, 2856–2868 (2017). doi: 10.1158/1078-0432.CCR-16-1114; pmid: 27986745

95. S. Goel *et al.*, Overcoming Therapeutic Resistance in HER2-Positive Breast Cancers with CDK4/6 Inhibitors. *Cancer Cell* **29**, 255–269 (2016). doi: 10.1016/j.ccell.2016.02.006; pmid: 26977878

96. J. L. Dean, A. K. McClendon, E. S. Knudsen, Modification of the DNA damage response by therapeutic CDK4/6 inhibition. *J. Biol. Chem.* **287**, 29075–29087 (2012). doi: 10.1074/jbc.M112.365494; pmid: 22733811

97. Y. Pikman *et al.*, Synergistic Drug Combinations with a CDK4/6 Inhibitor in T-cell Acute Lymphoblastic Leukemia. *Clin. Cancer Res.* **23**, 1012–1024 (2017). doi: 10.1158/1078-0432.CCR-15-2869; pmid: 28151717

98. J. Cao *et al.*, Combining CDK4/6 inhibition with taxanes enhances anti-tumor efficacy by sustained impairment of pRB-E2F pathways in squamous cell lung cancer. *Oncogene* **38**, 4125–4141 (2019). doi: 10.1038/s41388-019-0708-7; pmid: 30700828

99. Y. Gao *et al.*, Inhibition of CDK4 sensitizes multidrug resistant ovarian cancer cells to paclitaxel by increasing apoptosis. *Cell Oncol.* **40**, 209–218 (2017). doi: 10.1007/s13402-017-0316-x; pmid: 28243976

100. B. Salvador-Barbero *et al.*, CDK4/6 Inhibitors Impair Recovery from Cytotoxic Chemotherapy in Pancreatic Adenocarcinoma. *Cancer Cell* **37**, 340–353.e6 (2020). doi: 10.1016/j.ccell.2020.01.007; pmid: 32109375

101. J. Yi *et al.*, MYC status as a determinant of synergistic response to Olaparib and Palbociclib in ovarian cancer. *EBioMedicine* **43**, 225–237 (2019). doi: 10.1016/j.ebiom.2019.03.027; pmid: 30898650

102. D. Tempka *et al.*, Downregulation of PARP1 transcription by CDK4/6 inhibitors sensitizes human lung cancer cells to anticancer drug-induced death by impairing OGG1-dependent base excision repair. *Redox Biol.* **15**, 316–326 (2018). doi: 10.1016/j.redox.2017.12.017; pmid: 29306194

103. S. M. Johnson *et al.*, Mitigation of hematologic radiation toxicity in mice through pharmacological quiescence induced by CDK4/6 inhibition. *J. Clin. Invest.* **120**, 2528–2536 (2010). doi: 10.1172/JCI41402; pmid: 20577054

104. S. He *et al.*, Transient CDK4/6 inhibition protects hematopoietic stem cells from chemotherapy-induced exhaustion. *Sci. Transl. Med.* **9**, eaal3986 (2017). doi: 10.1126/scitranslmed.aal3986; pmid: 28446688

105. J. M. Weiss *et al.*, Myelopreservation with the CDK4/6 inhibitor trilaciclib in patients with small-cell lung cancer receiving first-line chemotherapy: A phase Ib/randomized phase II trial. *Ann. Oncol.* **30**, 1613–1621 (2019). doi: 10.1093/annonc/mdz278; pmid: 31504118

106. A. R. Tan *et al.*, Trilaciclib plus chemotherapy versus chemotherapy alone in patients with metastatic triple-negative breast cancer: A multicentre, randomised, open-label, phase 2 trial. *Lancet Oncol.* **20**, 1587–1601 (2019). doi: 10.1016/S1470-2045(19)30616-3; pmid: 31575503

107. J. O'Shaughnessy *et al.*, "Trilaciclib improves overall survival when given with gemcitabine/carboplatin in patients with metastatic triple-negative breast cancer." Paper presented at the 2020 San Antonio Breast Cancer Symposium, abstract PD1-06 (2020).

108. J. Franco, U. Balaji, E. Freinkman, A. K. Witkiewicz, E. S. Knudsen, Metabolic Reprogramming of Pancreatic Cancer Mediated by CDK4/6 Inhibition Elicits Unique Vulnerabilities. *Cell Rep.* **14**, 979–990 (2016). doi: 10.1016/j.celrep.2015.12.094; pmid: 26804906

109. Q. Yin *et al.*, CDK4/6 regulate lysosome biogenesis through TFEB/TFE3. *J. Cell Biol.* **219**, e201911036 (2020). doi: 10.1083/jcb.201911036; pmid: 32662822

110. L. Martínez-Carreres *et al.*, CDK4 Regulates Lysosomal Function and mTORC1 Activation to Promote Cancer Cell Survival. *Cancer Res.* **79**, 5245–5259 (2019). doi: 10.1158/0008-5472.CAN-19-0708; pmid: 31395606

111. A. Fassl *et al.*, Increased lysosomal biomass is responsible for the resistance of triple-negative breast cancers to CDK4/6 inhibition. *Sci. Adv.* **6**, eabb2210 (2020). doi: 10.1126/sciadv.abb2210; pmid: 32704543

112. H. Wang *et al.*, The metabolic function of cyclin D3-CDK6 kinase in cancer cell survival. *Nature* **546**, 426–430 (2017). doi: 10.1038/nature22797; pmid: 28607489

113. C. Luo *et al.*, Obesity/Type 2 Diabetes-Associated Liver Tumors Are Sensitive to Cyclin D1 Deficiency. *Cancer Res.* **80**, 3215–3221 (2020). doi: 10.1158/0008-5472.CAN-20-0106; pmid: 32606000

114. S. Goel *et al.*, CDK4/6 inhibition triggers anti-tumour immunity. *Nature* **548**, 471–475 (2017). doi: 10.1038/nature23465; pmid: 28813415

115. J. Deng *et al.*, CDK4/6 Inhibition Augments Antitumor Immunity by Enhancing T-cell Activation. *Cancer Discov.* **8**, 216–233 (2018). doi: 10.1158/2159-8290.CD-17-0915; pmid: 29101163

116. D. A. Schaer *et al.*, The CDK4/6 Inhibitor Abemaciclib Induces a T Cell Inflamed Tumor Microenvironment and Enhances the Efficacy of PD-L1 Checkpoint Blockade. *Cell Rep.* **22**, 2978–2994 (2018). doi: 10.1016/j.celrep.2018.02.053; pmid: 29539425

117. J. Yu *et al.*, Genetic Aberrations in the CDK4 Pathway Are Associated with Innate Resistance to PD-1 Blockade in Chinese Patients with Non-Cutaneous Melanoma. *Clin. Cancer Res.* **25**, 6511–6523 (2019). doi: 10.1158/1078-0432.CCR-19-0475; pmid: 31375512

118. J. Zhang *et al.*, Cyclin D-CDK4 kinase destabilizes PD-L1 via cullin 3-SPOP to control cancer immune surveillance. *Nature* **553**, 91–95 (2018). doi: 10.1038/nature25015; pmid: 29160310

119. X. Jin *et al.*, Phosphorylated RB Promotes Cancer Immunity by Inhibiting NF-κB Activation and PD-L1 Expression. *Mol. Cell* **73**, 22–35.e6 (2019). doi: 10.1016/j.molcel.2018.10.034; pmid: 30527665

120. L. Jerby-Arnon *et al.*, A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell* **175**, 984–997.e24 (2018). doi: 10.1016/j.cell.2018.09.006; pmid: 30388455

121. Z. L. Teo *et al.*, Combined CDK4/6 and PI3Kα Inhibition Is Synergistic and Immunogenic in Triple-Negative Breast Cancer. *Cancer Res.* **77**, 6340–6352 (2017). doi: 10.1158/0008-5472.CAN-17-2210; pmid: 28947417

122. R. S. Finn *et al.*, Palbociclib and Letrozole in Advanced Breast Cancer. *N. Engl. J. Med.* **375**, 1925–1936 (2016). doi: 10.1056/NEJMoa1607303; pmid: 27959613

123. N. C. Turner *et al.*, Palbociclib in Hormone-Receptor-Positive Advanced Breast Cancer. *N. Engl. J. Med.* **373**, 209–219 (2015). doi: 10.1056/NEJMoa1505270; pmid: 26030518

124. M. Cristofanilli *et al.*, Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): Final analysis of the multicentre, double-blind, phase 3 randomised controlled trial. *Lancet Oncol.* **17**, 425–439 (2016). doi: 10.1016/S1470-2045(15)00613-0; pmid: 26947331

125. G. N. Hortobagyi *et al.*, Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. *N. Engl. J. Med.* **375**, 1738–1748 (2016). doi: 10.1056/NEJMoa1609709; pmid: 27717303

126. D. J. Slamon *et al.*, Phase III Randomized Study of Ribociclib and Fulvestrant in Hormone Receptor-Positive, Human Epidermal Growth Factor Receptor 2-Negative Advanced Breast Cancer: MONALEESA-3. *J. Clin. Oncol.* **36**, 2465–2472 (2018). doi: 10.1200/JCO.2018.78.9909; pmid: 29860922

127. D. Tripathy *et al.*, Ribociclib plus endocrine therapy for premenopausal women with hormone-receptor-positive, advanced breast cancer (MONALEESA-7): A randomised phase 3 trial. *Lancet Oncol.* **19**, 904–915 (2018). doi: 10.1016/S1470-2045(18)30292-4; pmid: 29804902

128. G. W. Sledge Jr *et al.*, MONARCH 2: Abemaciclib in Combination With Fulvestrant in Women With HR+/HER2-Advanced Breast Cancer Who Had Progressed While Receiving Endocrine Therapy. *J. Clin. Oncol.* **35**, 2875–2884 (2017). doi: 10.1200/JCO.2017.73.7585; pmid: 28580882

129. M. P. Goetz *et al.*, MONARCH 3: Abemaciclib As Initial Therapy for Advanced Breast Cancer. *J. Clin. Oncol.* **35**, 3638–3646 (2017). doi: 10.1200/JCO.2017.75.6155; pmid: 28968163

130. S. Johnston *et al.*, MONARCH 3 final PFS: A randomized study of abemaciclib as initial therapy for advanced breast cancer. *NPJ Breast Cancer* **5**, 5 (2019). doi: 10.1038/s41523-018-0097-z; pmid: 30675515

131. S. A. Im *et al.*, Overall Survival with Ribociclib plus Endocrine Therapy in Breast Cancer. *N. Engl. J. Med.* **381**, 307–316 (2019). doi: 10.1056/NEJMoa1903765; pmid: 31166679

132. D. J. Slamon *et al.*, Overall Survival with Ribociclib plus Fulvestrant in Advanced Breast Cancer. *N. Engl. J. Med.* **382**, 514–524 (2020). doi: 10.1056/NEJMoa1911149; pmid: 31826360

133. G. W. Sledge Jr *et al.*, The Effect of Abemaciclib Plus Fulvestrant on Overall Survival in Hormone Receptor-Positive, ERBB2-Negative Breast Cancer That Progressed on Endocrine Therapy-MONARCH 2: A Randomized Clinical Trial. *JAMA Oncol.* **6**, 116–124 (2020). pmid: 31563959

134. N. C. Turner *et al.*, Overall Survival with Palbociclib and Fulvestrant in Advanced Breast Cancer. *N. Engl. J. Med.* **379**, 1926–1936 (2018). doi: 10.1056/NEJMoa1810527; pmid: 30345905

135. M. N. Dickler *et al.*, MONARCH 1, A Phase II Study of Abemaciclib, a CDK4 and CDK6 Inhibitor, as a Single Agent, in Patients with Refractory HR⁺/HER2⁻ Metastatic Breast Cancer. *Clin. Cancer Res.* **23**, 5218–5224 (2017). doi: 10.1158/1078-0432.CCR-17-0754; pmid: 28533223

136. S. R. D. Johnston *et al.*, Abemaciclib Combined With Endocrine Therapy for the Adjuvant Treatment of HR+, HER2-, Node-Positive, High-Risk, Early Breast Cancer (monarchE). *J. Clin. Oncol.* **38**, 3987–3998 (2020). doi: 10.1200/JCO.20.02514; pmid: 32954927

137. E. L. Mayer *et al.*, LBA12 - PALLAS: A randomized phase III trial of adjuvant palbociclib with endocrine therapy versus endocrine therapy alone for HR+/HER2- early breast cancer. *Ann. Oncol.* **31** (suppl. 4), S1142–S1215 (2020). doi: 10.1016/j.annonc.2020.08.2240

138. S. Pernas, S. M. Tolaney, E. P. Winer, S. Goel, CDK4/6 inhibition in breast cancer: Current practice and future directions. *Ther. Adv. Med. Oncol.* **10**, 1758835918786451 (2018). doi: 10.1177/1758835918786451; pmid: 30038670

139. www.clinicaltrials.gov/.

140. J. L. Dean *et al.*, Therapeutic response to CDK4/6 inhibition in breast cancer defined by ex vivo analyses of human tumors. *Cell Cycle* **11**, 2756–2761 (2012). doi: 10.4161/cc.21195; pmid: 22767154

141. M. T. Herrera-Abreu *et al.*, Early Adaptation and Acquired Resistance to CDK4/6 Inhibition in Estrogen Receptor-Positive Breast Cancer. *Cancer Res.* **76**, 2301–2313 (2016). doi: 10.1158/0008-5472.CAN-15-0728; pmid: 27020857

142. R. Condorelli *et al.*, Polyclonal RB1 mutations and acquired resistance to CDK 4/6 inhibitors in patients with metastatic breast cancer. *Ann. Oncol.* **29**, 640–645 (2018). doi: 10.1093/annonc/mdx784; pmid: 29236940

143. B. O'Leary *et al.*, The Genetic Landscape and Clonal Evolution of Breast Cancer Resistance to Palbociclib plus Fulvestrant in the PALOMA-3 Trial. *Cancer Discov.* **8**, 1390–1403 (2018). doi: 10.1158/2159-8290.CD-18-0264; pmid: 30206110

144. C. Costa *et al.*, PTEN Loss Mediates Clinical Cross-Resistance to CDK4/6 and PI3Kα Inhibitors in Breast Cancer. *Cancer Discov.* **10**, 72–85 (2020). doi: 10.1158/2159-8290.CD-18-0830; pmid: 31594766

145. S. A. Wander *et al.*, The Genomic Landscape of Intrinsic and Acquired Resistance to Cyclin-Dependent Kinase 4/6 Inhibitors in Patients with Hormone Receptor-Positive Metastatic Breast Cancer. *Cancer Discov.* **10**, 1174–1193 (2020). doi: 10.1158/2159-8290.CD-19-1390; pmid: 32404308

146. C. Yang *et al.*, Acquired CDK6 amplification promotes breast cancer resistance to CDK4/6 inhibitors and loss of ER signaling and dependence. *Oncogene* **36**, 2255–2264 (2017). doi: 10.1038/onc.2016.379; pmid: 27748766

147. L. Formisano *et al.*, Aberrant FGFR signaling mediates resistance to CDK4/6 inhibitors in ER+ breast cancer. *Nat. Commun.* **10**, 1373 (2019). doi: 10.1038/s41467-019-09068-2; pmid: 30914635

148. M. Álvarez-Fernández, M. Malumbres, Mechanisms of Sensitivity and Resistance to CDK4/6 Inhibition. *Cancer Cell* **37**, 514–529 (2020). doi: 10.1016/j.ccell.2020.03.010; pmid: 32289274

149. L. Cen *et al.*, p16-Cdk4-Rb axis controls sensitivity to a cyclin-dependent kinase inhibitor PD0332991 in glioblastoma xenograft cells. *Neuro-oncol.* **14**, 870–881 (2012). doi: 10.1093/neuonc/nos114; pmid: 22711607

150. M. E. Olanich *et al.*, CDK4 Amplification Reduces Sensitivity to CDK4/6 Inhibition in Fusion-Positive Rhabdomyosarcoma. *Clin. Cancer Res.* **21**, 4947–4959 (2015). doi: 10.1158/1078-0432.CCR-14-2955; pmid: 25810375

151. L. Cornell, S. A. Wander, T. Visal, N. Wagle, G. I. Shapiro, MicroRNA-Mediated Suppression of the TGF-β Pathway Confers Transmissible and Reversible CDK4/6 Inhibitor Resistance. *Cell Rep.* **26**, 2667–2680.e7 (2019). doi: 10.1016/j.celrep.2019.02.023; pmid: 30840889

152. R. de Leeuw *et al.*, MAPK Reliance via Acquired CDK4/6 Inhibitor Resistance in Cancer. *Clin. Cancer Res.* **24**, 4201–4214 (2018). doi: 10.1158/1078-0432.CCR-18-0410; pmid: 29739788

153. E. Haines *et al.*, Palbociclib resistance confers dependence on an FGFR-MAP kinase-mTOR-driven pathway in *KRAS*-mutant non-small cell lung cancer. *Oncotarget* **9**, 31572–31589 (2018). doi: 10.18632/oncotarget.25803; pmid: 30167080

154. G. Romano *et al.*, A Preexisting Rare PIK3CA^E545K Subpopulation Confers Clinical Resistance to MEK plus CDK4/6 Inhibition in NRAS Melanoma and Is Dependent on S6K1 Signaling. *Cancer Discov.* **8**, 556–567 (2018). doi: 10.1158/2159-8290.CD-17-0745; pmid: 29496665

155. D. B. Bowe, N. J. Kenney, Y. Adereth, I. G. Maroulakou, Suppression of Neu-induced mammary tumor growth in cyclin D1 deficient mice is compensated for by cyclin E. *Oncogene* **21**, 291–298 (2002). doi: 10.1038/sj.onc.1205025; pmid: 11803472

156. Y. Geng *et al.*, Rescue of cyclin D1 deficiency by knockin cyclin E. *Cell* **97**, 767–777 (1999). doi: 10.1016/S0092-8674(00)80788-6; pmid: 10380928

157. L. Wang *et al.*, Pharmacologic inhibition of CDK4/6: Mechanistic evidence for selective activity or acquired resistance in acute myeloid leukemia. *Blood* **110**, 2075–2083 (2007). doi: 10.1182/blood-2007-02-071266; pmid: 17537993

158. A. E. Vilgelm *et al.*, MDM2 antagonists overcome intrinsic resistance to CDK4/6 inhibition by inducing p21. *Sci. Transl. Med.* **11**, eaav7171 (2019). doi: 10.1126/scitranslmed.aav7171; pmid: 31413145

159. E. S. Knudsen *et al.*, Cell cycle plasticity driven by MTOR signaling: Integral resistance to CDK4/6 inhibition in patient-derived models of pancreatic cancer. *Oncogene* **38**, 3355–3370 (2019). doi: 10.1038/s41388-018-0650-0; pmid: 30696953

160. V. M. Jansen *et al.*, Kinome-Wide RNA Interference Screen Reveals a Role for PDK1 in Acquired Resistance to CDK4/6 Inhibition in ER-Positive Breast Cancer. *Cancer Res.* **77**, 2488–2499 (2017). doi: 10.1158/0008-5472.CAN-16-2653; pmid: 28249908

161. A. Yoshida *et al.*, SLC36A1-mTORC1 signaling drives acquired resistance to CDK4/6 inhibitors. *Sci. Adv.* **5**, eaax6352 (2019). doi: 10.1126/sciadv.aax6352; pmid: 31555743

162. J. I. Griffiths *et al.*, Convergent evolution of resistance pathways during early stage breast cancer treatment with combination cell cycle (CDK) and endocrine inhibitors. *Cancer Res.* **81** (suppl. 4), SP012 (2021). doi: 10.1101/2021.01.19.427299

163. R. L. Hoffman, "The discovery of PF-07104091: A CDK2 selective inhibitor for the treatment of cyclinE amplified cancers." Paper presented at the AACR Annual Meeting 2021, Session DDT02—New Drugs on the Horizon: Part 3 (2021).

164. Y. J. Choi *et al.*, Development of a selective CDK2-E inhibitor in CCNE driven cancers. *Cancer Res.* **81** (suppl. 13), 1279 (2021). doi: 10.1158/1538-7445.AM2021-1279

165. K. Freeman-Cook *et al.*, Expanding control of the tumor cell cycle with a CDK2/4/6 inhibitor. *Cancer Cell* **39**, 1404–1421.e11 (2021). doi: 10.1016/j.ccell.2021.08.009; pmid: 34520734

166. P. Hydbring, M. Malumbres, P. Sicinski, Non-canonical functions of cell cycle cyclins and cyclin-dependent kinases. *Nat. Rev. Mol. Cell Biol.* **17**, 280–292 (2016). doi: 10.1038/nrm.2016.27; pmid: 27033256
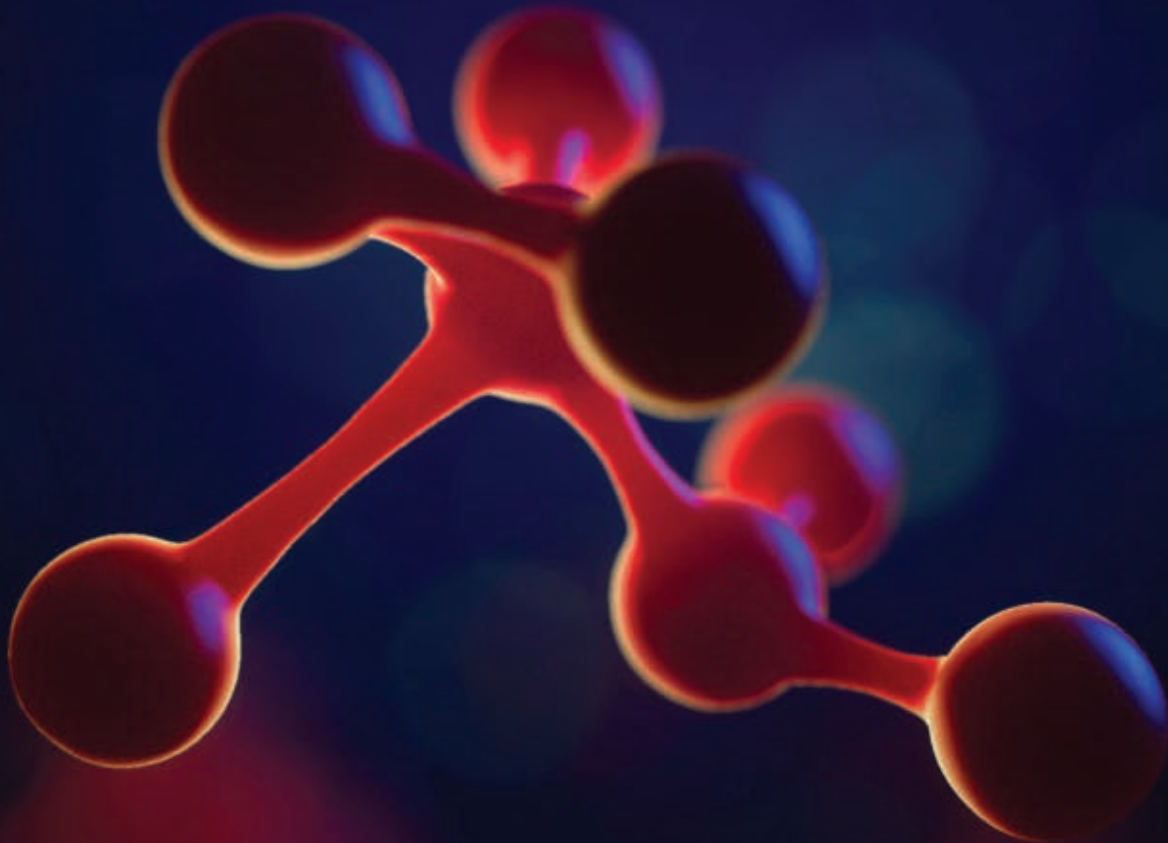
CELL BIOLOGY

Mitochondria shed their outer membrane in response to infection-induced stress

Xianhe Li, Julian Straub, Tânia Catarina Medeiros, Chahat Mehra, Fabian den Brave, Esra Peker, Ilian Atanassov, Katharina Stillger, Jonas Benjamin Michaelis, Emma Burbridge, Colin Adrain, Christian Münch, Jan Riemer, Thomas Becker, Lena F. Pernas*

INTRODUCTION: Mitochondria are dynamic organelles that coordinate many cellular functions that are essential for life, including diverse metabolic processes, cell division and differentiation, and immune signaling.

To carry out these diverse functions, mitochondria must communicate with the cytosol, a task mediated by the outer mitochondrial membrane (OMM), the gateway between mitochondria and the rest of the cell. Thus, preserving the integrity of the OMM is essential for cellular homeostasis. Although responses to stress that is artificially induced by small molecules have been described, little is known of the mechanisms by which mammalian cells respond to naturally occurring stresses of the OMM.

RATIONALE: Several intracellular microbes come in contact with the host OMM or release proteins that target the OMM. We reasoned that microbial infection would serve as a model with which to study cellular responses to natural OMM stress. To this end, we chose the human parasite *Toxoplasma gondii* because it tethers host mitochondria to its parasite vacuole; in an infected cell, areas of close membrane apposition form between the host OMM and the parasite vacuole membrane. To address how *Toxoplasma* affected the OMM, we performed live-cell imaging of infected mammalian cells stably expressing OMM-targeted green fluorescent protein (GFP). We found that mitochondria in contact with the *Toxoplasma* vacuole released large structures that were positive for OMM, which we termed "SPOTs." Analysis of SPOTs in fixed and live cells revealed that SPOTs did not contain proteins of the mitochondrial matrix and inner mitochondrial membrane (IMM).

RESULTS: Having identified *Toxoplasma* infection as a natural stress that induced OMM remodeling and the shedding of SPOTs, we next dissected how these structures are formed. We found that the secreted effector protein TgMAF1 (*Toxoplasma gondii* mitochondrial association factor 1), which tethers the host OMM to the parasite vacuole membrane, was required for SPOT formation. TgMAF1 led to a decrease in the amount of mitochondrial proteins during infection. In particular, the OMM proteins mitofusin 1 and mitofusin 2 were degraded during infection. These proteins, which mediate a nutritional defense against *Toxoplasma* by promoting mitochondrial uptake of fatty acids, localized to SPOTs. The ability of TgMAF1 to induce SPOTs depended on its binding to the host OMM import receptor TOM70 (translocase of the outer membrane 70), whose import function

TgMAF1 impaired. TOM70 was required for optimal parasite growth and enabled an interaction between TgMAF1 and the OMM translocase SAM50 (sorting assembly machinery 50 kDa subunit), which is a key component of the OMM-IMM mitochondrial intermembrane space bridging (MIB) complex. The genetic ablation of SAM50 and the overexpression of an OMM-targeted protein induced the formation of SPOT-like structures independently of infection.

CONCLUSION: Because SAM50 is the only component of mitochondrial import machinery that bridges the OMM and IMM, it is in a position to translate OMM stress into the removal of compromised OMM. TgMAF1 behaves as a mitochondrial preprotein that uses the host receptor TOM70 to interact with SAM50. This enables *Toxoplasma* to hijack a cellular response to OMM stress—the formation of SPOTs—and drive the constitutive shedding of the OMM. Consequently, levels of mitochondrial proteins that restrict parasite growth are depleted, and import machinery that is required for mitochondrial biogenesis is sequestered on SPOTs. In an infection-independent context, however, we propose that SPOT-like structures could mitigate OMM stress by enabling the excision of dysfunctional OMM machinery, such as import receptors or translocases during defective import. Our finding that OMM remodeling occurs during infection and infection-independent scenarios sheds light on potential cellular mechanisms that safeguard OMM function. ∎

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abi4343
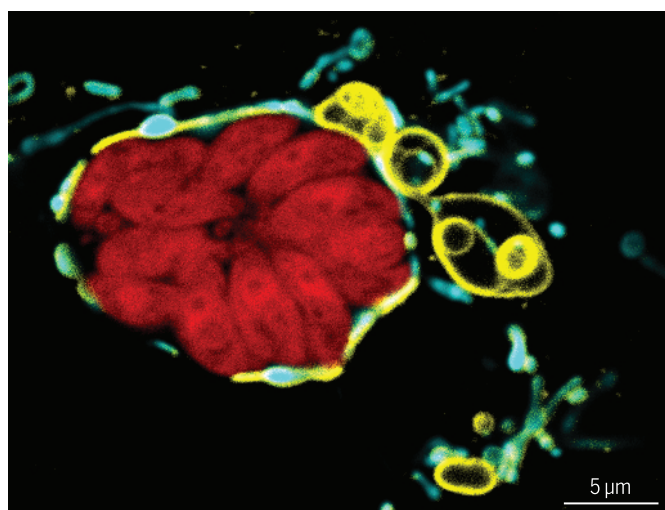
Mitochondria shed SPOTs during *Toxoplasma* infection. (**Left**) Mitochondria surrounding the parasite *Toxoplasma* (red) shed SPOTs, large structures positive for OMM (yellow) that lack mitochondrial matrix (cyan). (**Right**) Cartoon depiction of image at left as infection progresses (clock). OMM proteins—including the import translocase SAM50, and MFN1 and MFN2, which are required for mitochondrial fusion—are sequestered on SPOTs.
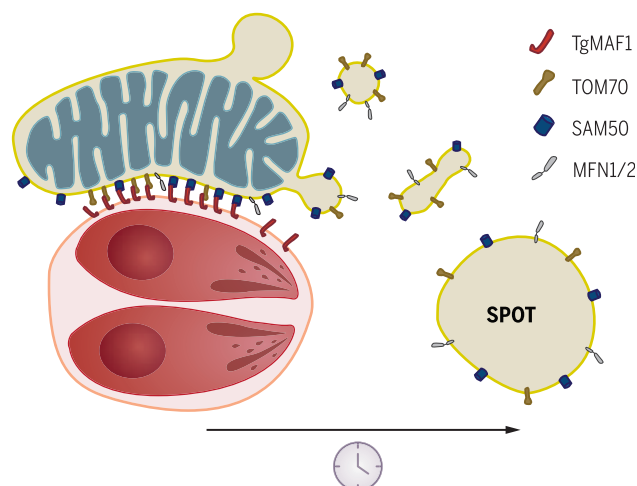
CELL BIOLOGY

Mitochondria shed their outer membrane in response to infection-induced stress

Xianhe Li[1], Julian Straub[1], Tânia Catarina Medeiros[1], Chahat Mehra[1], Fabian den Brave[2], Esra Peker[3], Ilian Atanassov[1], Katharina Stillger[3], Jonas Benjamin Michaelis[4], Emma Burbridge[5,6], Colin Adrain[5,6], Christian Münch[4], Jan Riemer[3,7], Thomas Becker[2], Lena F. Pernas[1,7]*

The outer mitochondrial membrane (OMM) is essential for cellular homeostasis. Yet little is known of the mechanisms that remodel it during natural stresses. We found that large "SPOTs" (structures positive for OMM) emerge during *Toxoplasma gondii* infection in mammalian cells. SPOTs mediated the depletion of the OMM proteins mitofusin 1 and 2, which restrict parasite growth. The formation of SPOTs depended on the parasite effector TgMAF1 and the host mitochondrial import receptor TOM70, which is required for optimal parasite proliferation. TOM70 enabled TgMAF1 to interact with the host OMM translocase SAM50. The ablation of SAM50 or the overexpression of an OMM-targeted protein promoted OMM remodeling independently of infection. Thus, *Toxoplasma* hijacks the formation of SPOTs, a cellular response to OMM stress, to promote its growth.

Mitochondria are dynamic organelles that coordinate cellular functions essential for life, including diverse metabolic processes, cell division and differentiation, and immune signaling (*1*). At the interface between mitochondria and the rest of the cell, the outer mitochondrial membrane (OMM) plays a central role in all mitochondrial functions. Arguably, the most vital of these functions is to enable mitochondrial biogenesis by providing a platform for the machinery that imports almost all ~1200 nuclear-encoded mitochondrial proteins in mammals (*2*). Dysfunction of the mitochondrial import machinery results in systemic pathology of the organelle and organismal death (*3*). Because all mitochondrial functions rely on the import of proteins into the organelle, quality control pathways must regulate import processes and prevent the toxic accumulation of nonimported or mislocalized precursor proteins, several of which have been identified and studied in yeast (*4*). By contrast, how the OMM is protected and remodeled during import stress in mammals is largely uncharacterized.

Intracellular microbes pose a naturally occurring threat to mitochondria, which is perhaps the evolutionary consequence of the role that these organelles play in antimicrobial immune signaling and nutritional defense (*5*–*8*). Several bacteria, for example, release effector proteins that localize to the OMM and perturb mitochondrial function (*9*). We used the parasite *Toxoplasma gondii*, which infects up to one-third of the human population and has an unparalleled host range, to investigate the cellular responses that remodel the OMM (*10*). With respect to the OMM, *Toxoplasma* is of particular interest because it is one of several pathogens to reside in vacuoles found in close physical proximity to mitochondria (*11*, *12*).

Mitochondria shed SPOTs during *Toxoplasma* infection

To monitor the impact of *Toxoplasma* infection on the OMM, we infected mouse cells stably expressing enhanced green fluorescent protein (eGFP) fused to the OMM-targeting transmembrane (TM) domain of OMP25 (OMM-GFP) with mCherry-expressing *Toxoplasma* (*13*). As early as 6 hours after infection, we observed large spherical and elliptical structures enriched for OMM-GFP that were absent in uninfected cells, which we termed "SPOTs" (structures positive for outer mitochondrial membrane) (Fig. 1A). The percentage of SPOT-positive cells increased over the course of infection (Fig. 1B). At 24 hours after infection, >50% of infected cells contained between 1 and 20 SPOTs that were on average ~2.6 μm in diameter (Fig. 1, A to D). To exclude the possibility that the formation of SPOTs depended on OMM-GFP expression, we also examined infected cells using a fluorescent phosphatidylcholine conjugate (FL-HPC) that in part integrates into mitochondrial membranes. In both wild-type (WT) and OMM–BFP (blue fluorescent protein)–expressing U2OS (human) cells, FL-HPC distributed to SPOT-like structures

that were only present in infected cells (fig. S1). Thus, SPOT formation is a general consequence of *Toxoplasma* infection in mammalian cells.

Despite their large size, we suspect that SPOTs have been overlooked because although they retain OMM proteins such as TOM20 (translocase of the outer membrane 20), they are not labeled by mitotracker, a commonly used vital dye that accumulates in mitochondria in a membrane potential–dependent manner (Fig. 1, E to F, and figs. S1 and S2A). Furthermore, in live-cell imaging experiments, SPOTs also lack fluorescent markers of other mitochondrial compartments, including the matrix and inner mitochondrial membrane (IMM) (Fig. 1, G to J). We confirmed that SPOTs also lacked endogenous levels of proteins of the intermembrane space (IMS) (AIFM1), IMM (ATPF1B), and matrix (CS and mtHSP70) (fig. S2, B to E). Thus, the OMM is remodeled to release SPOTs during *Toxoplasma* infection. Hereafter, we refer to SPOTs as structures that are positive for an OMM marker but that lack markers of other mitochondrial compartments.

The large diameter and lack of mitochondrial matrix proteins in SPOTs suggested that they differed from fragmented mitochondria as well as mitochondria-derived compartments (MDCs) and mitochondria-derived vesicles (MDVs), which form in response to amino acid toxicity and oxidative stress, respectively (*14*–*19*). MDCs average 1 μm in diameter and depend on the OMM guanosine triphosphatase (GTPase) MIRO1 (mitochondrial rho GTPase 1), which mediates mitochondrial trafficking (*16*, *17*). The smaller MDVs that transport cargo between mitochondria and peroxisomes or lysosomes—the latter of which depend on the E3 ligase parkin and the mitochondrial kinase PINK1 (PTEN-induced putative kinase protein 1)—range between 75 and 150 nm in diameter (*14*, *18*, *20*). Neither the loss of the key mitochondrial fission factor DRP1 (dynamin-related protein 1) nor of MIRO1 and its paralog MIRO2 significantly affected SPOT formation during infection (fig. S3). Moreover, SPOTs also formed during infection in HeLa cells that were deficient for PINK1 in a manner similar to that of WT HeLa cells that lacked detectable parkin expression (fig. S4) (*21*).

The dynamin-binding partner SNX9 (sorting nexin 9) that remodels endocytic membranes mediates the emergence of a subset of MDVs that contain the matrix proteins OGDH (oxoglutarate dehydrogenase) and mtHSP70 (*22*–*24*). We thus asked whether SNX9 also participated in SPOT formation. In infected cells, the depletion of SNX9 prevented the formation of SPOTs (fig. S5, A to E). However, *Toxoplasma*-induced SPOTs lacked both mtHSP70 and OGDH, matrix proteins that are markers of SNX9-dependent MDVs (figs. S2E and S5F). Thus, SPOTs represent an independent class of structures that bud from the OMM.

[1]Max Planck Institute for Biology of Ageing, Cologne, Germany. [2]Institute of Biochemistry and Molecular Biology, Medical Faculty, University of Bonn, Bonn, Germany. [3]Institute of Biochemistry, University of Cologne, Cologne, Germany. [4]Institute of Biochemistry II, Faculty of Medicine, Goethe University, Frankfurt am Main, Germany. [5]Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, Belfast, Northern Ireland. [6]Instituto Gulbenkian de Ciência, Oeiras, Portugal. [7]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany.
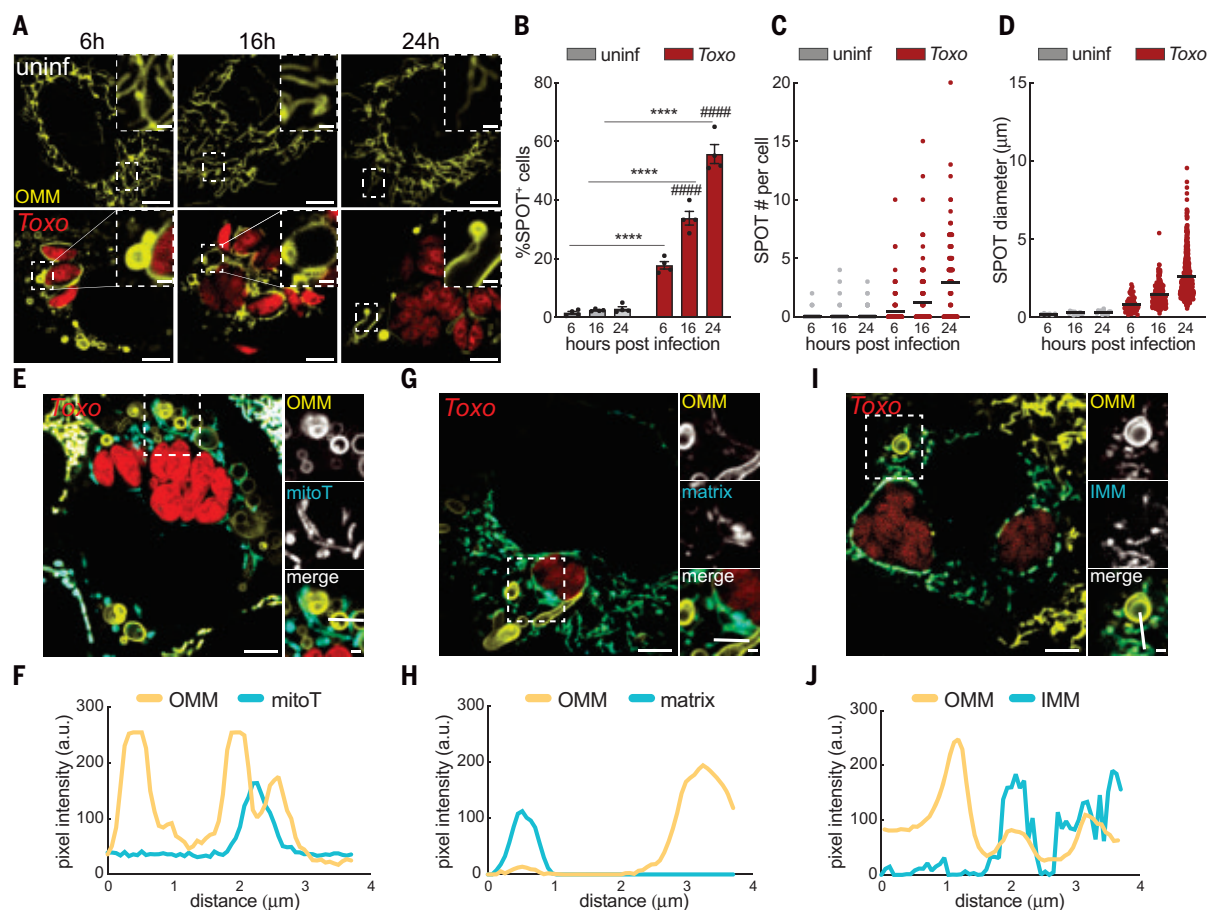*Corresponding author. Email: pernas@age.mpg.de

Fig. 1. The outer membrane is remodeled during *Toxoplasma* infection through the release of SPOTs. (**A**) Representative live-cell images of the OMM (GFP) in uninfected (uninf) and *Toxoplasma* (mCh)–infected (*Toxo*) MEFs at 6, 16, and 24 hours after infection. (**B**) Percentage of SPOT-positive cells in experiments as in (A); data are mean ± SEM of more than 100 cells counted from four biological replicates; ****$P < 0.0001$ for uninfected versus infected; ####$P < 0.0001$ for 6 hours after infection versus 16 and 24 hours after infection by means of two-way ANOVA. (**C** and **D**) Scatterplot with mean (C) number and (D) diameter of SPOTs in experiments as in (A) from more than 30 infected cells from three biological replicates. (**E**, **G**, and **I**) Representative live-cell images of the OMM (GFP) in *Toxoplasma* (mCherry)–infected MEFs (E) labeled with MitoTracker Deep Red (mitoT); (G) expressing matrix-targeted BFP (matrix); or (I) expressing RFP-tagged TIM50 (IMM). (**F**, **H**, and **J**) Corresponding pixel intensity plots for white line in (E), (G), and (I) insets, respectively. Scale bars, 5 µm and (insets) 1 µm.

The parasite effector TgMAF1 is required for SPOT formation and the depletion of host mitochondrial proteins

How does *Toxoplasma* induce the formation of SPOTs? In live-cell imaging analyses of infected cells, SPOTs were readily visualized emerging from mitochondria in proximity to the parasite vacuole (PV) (Figs. 1, A, E, G, and 2A and movies S1 and S2). At the mitochondria-*Toxoplasma* interface, the OMM is tethered to the PV membrane (PVM) at a distance of 12 to 17 nm, regions that can be referred to as "contact sites," a term that describes areas of close membrane apposition between organelles (*25*, *26*). Mitochondria-*Toxoplasma* PV contact sites, previously termed host mitochondrial association (HMA), are completely dependent on the parasite effector protein TgMAF1 (*Toxoplasma gondii* mitochondrial association factor 1) (*27*). TgMAF1 is anchored in the PVM and posited to interact with host mitochondria through its cytosolically exposed

C terminus (*27*–*29*). TgMAF1 affects the host inflammatory responses during infection, but its function at mitochondria-*Toxoplasma* contact sites remains elusive (*27*). To test the hypothesis that TgMAF1 tethering of host mitochondria contributes to SPOT formation, we assessed the frequency of SPOTs in cells infected with WT or Δ*maf1* parasites. The ablation of TgMAF1 completely prevented the formation of SPOTs (Fig. 2, B to E). Thus, TgMAF1 is required for the formation of SPOTs during infection.

The physical remodeling of the OMM during *Toxoplasma* infection prompted us to ask whether the formation of SPOTs also led to changes in its composition. To address this, we used proteomics to compare the protein abundances between whole cell extracts and mitochondria immunopurified (mitoIP) from uninfected cells and cells infected with WT or Δ*maf1* parasites (*13*). Infection with WT parasites caused a global decrease in the whole

cell and mitoIP abundance of mitochondrial proteins that was partially prevented during infection with Δ*maf1* parasites (Fig. 3, A to D; fig. S6, A to D; and data files S1 and S2). Thus, TgMAF1, beyond its tethering function, is also required for the decrease in mitochondrial proteins during infection.

A targeted analysis of changes in the OMM compartment from which SPOTs are derived revealed TgMAF1-dependent decreases in several proteins that mediate antimicrobial defense, including MIRO1, MIRO2, mitofusin 1 (MFN1), and MFN2, results we confirmed with immunoblotting (Fig. 3, E to F) (*5*, *30*). The depletion of MFN1 and MFN2 during infection was particularly intriguing given that these OMM GTPases are essential for mitochondria to restrict parasite proliferation and raised the question of how TgMAF1 drove their decrease (*5*). Infection did not reduce *MFN1* and *MFN2* mRNA amounts (fig. S7A). Furthermore, neither the loss of PINK1, a key
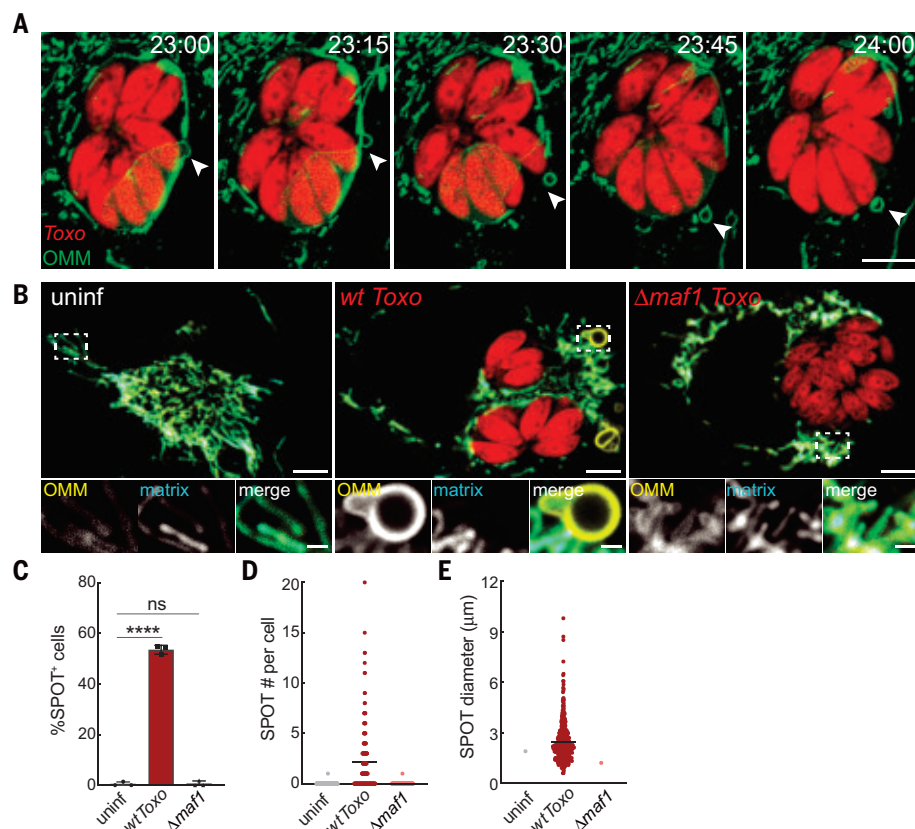
Fig. 2. TgMAF1 is required for SPOT formation. (**A**) Representative live-cell images of the OMM (GFP) in *Toxoplasma* (mCh)–infected HFFs. Shown is the formation of a SPOT (indicated with white arrowhead) over a 1-hour time-lapse movie starting at 23 hours after infection with frames captured every 15 min (movie S2). Scale bar, 5 µm. (**B**) Representative live-cell images of the OMM (GFP) and matrix (BFP) in uninf, WT *Toxo* (mCh)–infected, or Δ*maf1 Toxo* (mCh)–infected U2OS cells. Scale bars, 5 µm and (inset) 1 µm. (**C**) Percentage of SPOT-positive cells in experiments as in (B). Data are mean ± SEM of more than 100 cells counted from three biological replicates. ns, not significant; ****P < 0.0001 for uninfected versus infected by means of one-way ANOVA analysis. (**D** and **E**) Scatterplots with mean (D) number and (E) diameter of SPOTs in experiments as in (B) from more than 30 infected cells from three biological replicates.

mediator of mitophagy, nor the loss of the essential autophagy gene *Atg7* prevented the *Toxoplasma*-induced reduction in MFN1 and MFN2 (fig. S7, B and C). The decrease in MFN1 and MFN2 correlated with their redistribution from mitochondria to SPOTs that were also positive for FAF2 (Fas-associated factor family member 2), a protein that recruits the proteasomal machinery required for endoplasmic reticulum (ER)–associated degradation (Fig. 3, G to J, and fig. S7, D to F) (*31*). Because MFN1 and MFN2 are regulated by the ubiquitin/proteasome system (UPS) and FAF2 mediates the turnover of MFN1/2 (Fzo1) in yeast, we asked whether FAF2 played such a role in our system (*32*, *33*). In mouse embryonic fibroblasts (MEFs) in which FAF2 was ablated (FAF2 KO MEFs), the depletion of MFN1 and MFN2 in mitochondrial and whole-cell fractions that would normally be caused by infection was prevented (fig. S7G). Conversely, the expression of *Faf2* cDNA in FAF2 KO MEFs rescued their depletion during

infection (fig. S7H). To test a potential role for the UPS in mediating the degradation of MFN1 and MFN2, we turned to TAK-243, a small molecular inhibitor of the ubiquitin activation step that precedes degradation by the UPS because inhibitors of proteasomal machinery restrict parasite proliferation (fig. S7, I to J) (*34*). Treatment with TAK-243 did not affect parasite burden as assessed by *Toxoplasma* TgGAP45 but prevented the depletion of MFN1 and MFN2 and to a lesser extent that of MIRO1 and MIRO2, which also localized to SPOTs during infection (figs. S7, I to J, and S8). Thus, FAF2 mediates the proteolytic degradation of MFN1 and MFN2 on SPOTs, and certain OMM proteins that localize to SPOTs are targeted by the UPS for degradation.

TgMAF1 inhibits host TOM70 import function

How does TgMAF1 induce the remodeling of the OMM into SPOTs? We reasoned that identifying TgMAF1-interacting factor(s) might pro-

vide clues to host proteins involved in SPOT formation. To find these host factors, we immunopurified TgMAF1 from cells infected with Δ*maf1* parasites complemented with a hemagglutinin (HA)–tagged TgMAF1 (Δ*maf1:HA-MAF1*) at a multiplicity of infection (MOI) of 1, 2.4, and 6 and identified the major interaction partners using mass spectrometry. The OMM import receptor TOM70 was the most abundant host protein found at >256-fold enrichment in TgMAF1-IPs of cells infected at all MOIs relative to control IPs (Fig. 4A and data file S3). TOM70 is required for *Toxoplasma* association with host mitochondria and is enriched on mitochondria tethered to the PV (*35*). To confirm that parasite TgMAF1 and host TOM70 interact during infection, we also performed an immunoblot analysis of TgMAF1-IPs. TgMAF1 coimmunoprecipitated TOM70 but not other OMM proteins, including VDAC1 (voltage-dependent anion channel 1) or VDAC2, nor its putative interacting partner HSPA9 (heat shock protein family A member 9) (Fig. 4B) (*35*). Furthermore, stably expressed GFP-tagged TOM70 coimmunoprecipitated TgMAF1 from lysates of infected MEFs and was enriched on the OMM of a mitochondrion tethered to the PV at 20-fold higher concentrations than those of OMM regions of the same mitochondrion not in contact with the PV (Fig. 4, C to E). This consequence of infection was completely dependent on TgMAF1 because the distribution of TOM70 did not differ between cells infected with Δ*maf1:HA* parasites and uninfected cells (Fig. 4, D to E). To address whether TgMAF1 and TOM70 directly interact, we incubated TgMAF1 produced in a cell-free system with the purified cytosolic domain of yeast TOM70. Only full-length TgMAF1—but not a mutant lacking a predicted internal mitochondrial targeting sequence (iMTS), which mediates precursor binding to TOM70—bound to affinity-purified TOM70 (Fig. 4, F to G) (*35*, *36*). Thus, parasite TgMAF1 and host TOM70 directly interact during infection.

TOM70 has two critical functions at the OMM. First, TOM70 recruits cytosolic chaperones such as HSP90 that mediate precursor import and thus protects against proteotoxicity owing to accumulated precursors through its N-terminal CLAMP domain (*37*, *38*). Second, TOM70 is a major import receptor, most notably for SLC25 mitochondrial carriers that it binds through its CORE and C-tail domains (*38*). Thus, TgMAF1 could perturb TOM70 chaperone-binding activity or import activity to exploit a host response to import stress, leading to SPOT formation. To address these possibilities, we asked whether the abundances of HSP90 or SLC25 proteins were altered in our proteomics datasets of uninfected cells and cells infected with WT or Δ*maf1* parasites (Fig. 3, A to D). The levels of HSP90 were increased in mitoIPs in a
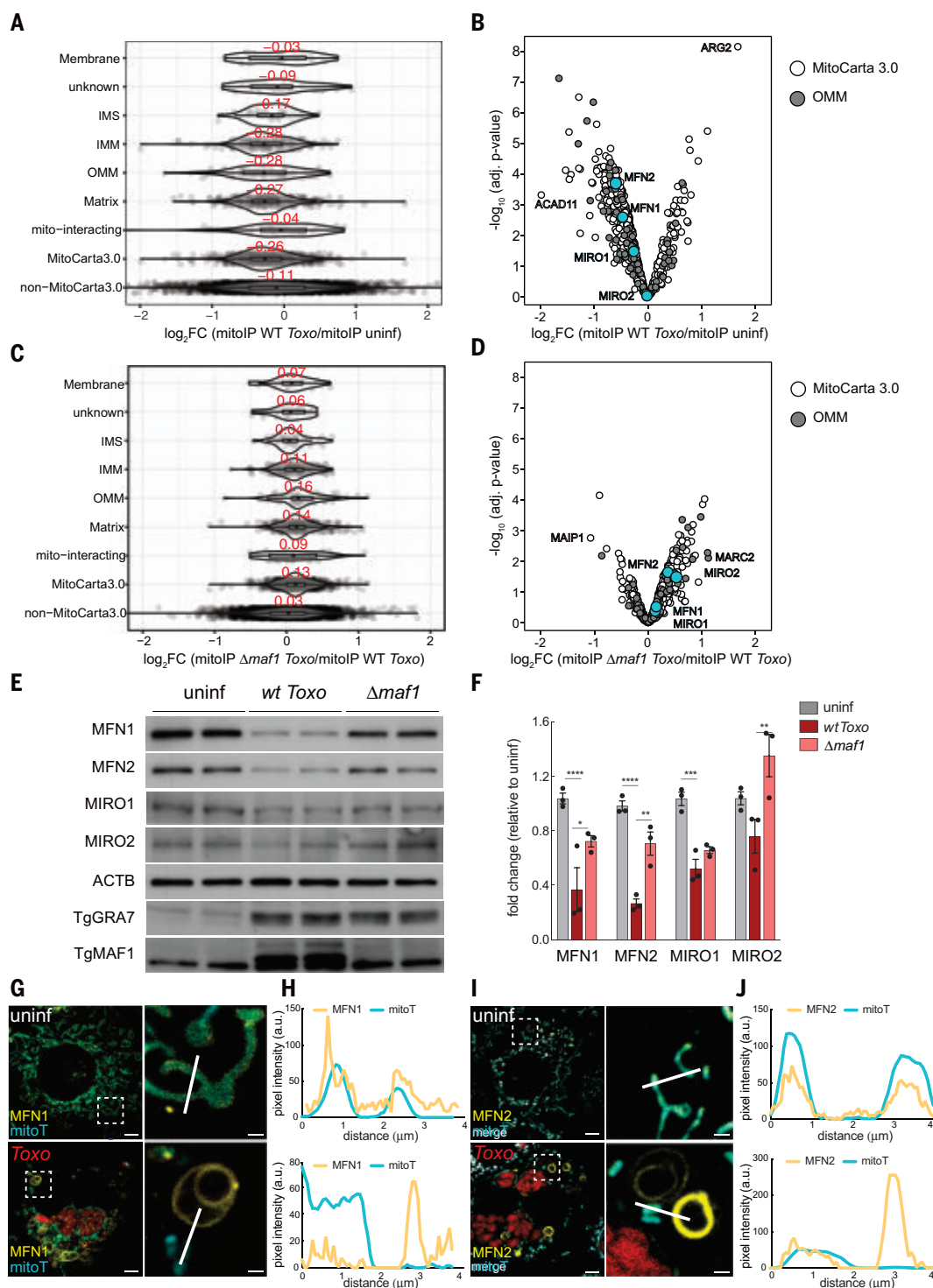
Fig. 3. TgMAF1 drives changes in the mitochondrial proteome and the depletion of certain OMM proteins that localize to SPOTs. (**A** and **C**) Mitochondria were immunopurified (mitoIP) from HeLa cells that were uninfected (uninf) ($n = 4$), WT Toxo-infected ($n = 4$), or Δmaf1 Toxo–infected ($n = 3$) at 24 hours after infection and analyzed by means of mass spectromtery. Shown are boxplots depicting the $\log_2$-fold change (FC) of detected MitoCarta3.0 proteins between indicated treatments according to submitochondrial localization. Median values for each subcompartment are indicated in red. (**B** and **D**) Volcano plot of proteins in (A) and (C). Dark gray, OMM proteins as classified by MitoCarta3.0; white, all other MitoCarta3.0 proteins. (**E**) Cells were uninf,

WT Toxo–, or Δmaf1-infected for 24 hours and analyzed by means of immunoblotting for MFN1, ~80 kda; MFN2, ~80 kDa; MIRO1, ~75 kDa; MIRO2, ~75 kDa; ACTB, ~45 kDa; TgGRA7, ~32 kDa; and TgMAF1, ~60 kDa. (**F**) Fold change in density of protein bands imaged as in (E) normalized to actin and relative to mean of uninfected samples. Data are mean ± SEM from three biological replicates; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ for WT Toxo versus uninfected or Δmaf1-infected samples. Representative live-cell images of (**G**) Mfn1$^{-/-}$:GFP-Mfn1 MEFs and (**I**) Mfn2$^{-/-}$ MEFs expressing mCh-MFN2 at 24 hours after infection with Toxoplasma (BFP). (**H** and **J**) Corresponding pixel intensity plots for white line in (G) and (I) insets, respectively. Scale bars, 5 μm and (inset) 1 μm.

TgMAF1-dependent manner (fig. S9, A and B). Opposite to HSP90 however, concentrations of 11 of the 22 detected SLC25 proteins were decreased in a MAF1-dependent manner, hinting at an impact of MAF1 on TOM70-dependent import (fig. S9, A to B). To more carefully dissect how TgMAF1 affects TOM70 import function, we turned to budding yeast, a workhorse for mitochondrial import studies and a system in which TOM70 function is well characterized (36, 37). After confirming that stably expressed TgMAF1 coimmunoprecipitated yeast TOM70, we assessed its effect on TOM70 import (fig. S9C). Mitochondria isolated from TgMAF1⁺ yeast imported the TOM70-dependent precursor AAC (ADP/ATP carrier) more slowly than did mitochondria from WT yeast, whereas the Su9–DHFR (dihydrofolate reductase) fusion protein, which does not require TOM70, was imported at similar rates (fig. S9, D to G) (37, 39). Furthermore, TgMAF1 stably expressed in mammalian cells coimmunoprecipitated TOM70 and impaired the import of AAC2 in a TOM70-dependent manner (Fig. 4, H to K). Thus, TgMAF1 negatively affects TOM70-dependent import.

Host TOM70 and SAM50 are required for infection-induced SPOT formation

To further explore a potential link between TOM70 inhibition and SPOT formation, we asked whether the expression of TgMAF1 alone—which was sufficient to impair TOM70-dependent import—also induced the formation of SPOTs. However, neither TgMAF1, which localizes to host mitochondria when expressed in mammalian cells, nor an ER-anchored TgMAF1 remodeled the OMM in a manner reminiscent of *Toxoplasma*-induced SPOTs, indicating a requirement for other parasite factors in their formation (fig. S10) (27). To test the impact of a more complete loss of TOM70 function, we generated cells lacking TOM70 (Fig. 5A). Contrary to our expectation, the loss of TOM70 did not induce SPOT formation but instead prevented their formation during infection (Fig. 5, B to E), whereas HSP90 inhibition with the specific inhibitor 17-DMAG (17-dimethylaminoethylamino-17-demethoxygeldanamycin) did not affect the rate of their formation (fig. S11). We also noticed that the loss of TOM70 impaired the growth of TgMAF1⁺ but not Δmaf1 parasites (Fig. 5F). Thus, we considered an alternate scenario in which *Toxoplasma* exploits host TOM70 to mediate the import of TgMAF1 into host mitochondria through an OMM translocase. Because TgMAF1 is anchored into the PVM, its stable interaction with OMM import machinery would be sensed as a chronic stress, activating a host response that remodels the OMM through the budding of SPOTs. The constant formation of SPOTs, however, would

lead to the shedding of OMM proteins, including import machinery and MFN1 and MFN2, which restrict parasite growth (Fig. 5G) (5). To test whether a TgMAF1-TOM70 interaction was critical for SPOT formation, we generated Δmaf1 parasites expressing a TgMAF1 mutant that exhibits a diminished interaction with TOM70 because of a trio of basic residues at its C terminus (TgMAF1ᴿᴷᴷ) (fig. S12, A to B) (35). Cells infected with Δmaf1:MAF1ᴿᴷᴷ mutant parasites were SPOT-less, despite retaining a significant capacity to tether mitochondria to the PVM (figs. S12, C to F, and S13, A and B). Thus, a stable TgMAF1-TOM70 interaction is essential for the formation of SPOTs.

Because the loss of TOM70 did not form SPOTs, we posited that the stress required for their formation was derived from a TOM70-dependent interaction between MAF1 and another OMM host import factor (Fig. 5G). In line with this possibility, a study of TgMAF1 interactors has identified SAM50 (sorting assembly machinery 50 kDa subunit), a translocase that is essential for the OMM integration of a subset of mitochondrial proteins, and MIC19 and MIC60, proteins with which SAM50 forms the OMM-IMM mitochondrial intermembrane space bridging (MIB) complex (29, 40). We therefore asked whether TOM70 was required for a TgMAF1-SAM50 interaction. SAM50 was present in TgMAF1-IPs from infected WT cells but not cells in which TOM70 was deleted (Fig. 6A). We obtained similar results for MIC19 and MIC60 (Fig. 6A). Thus, TOM70 is required for TgMAF1 to interact with the OMM translocase SAM50 and its interacting partners MIC60 and MIC19.

SAM50 is the only component of host mitochondrial import machinery with a defined role in bridging the OMM and IMM. We thus reasoned that TgMAF1 could induce the formation of SPOTs that are OMM-positive but lack IMM by exerting an effect on SAM50. SAM50, but not its MIB-complex partner MIC60 that is tethered to the IMM, was enriched on SPOTs in infected cells (Fig. 6B and fig. S14). To directly test the possibility that SPOTs resulted from a separation of the OMM and IMM, we used cell lines in which the MIB components *SAM50* and *MIC60* as well as *TOM70* were silenced with doxycycline-inducible short hairpin RNA (shRNA). We confirmed that TOM70 is required for SPOT formation (Fig. 6, C to F). The depletion of *SAM50* was sufficient to induce smaller SPOT-like structures in >25% of uninfected cells relative to <1% of cells expressing a control shRNA (Fig. 6, C to F). Infection-induced SPOT formation in *SAM50*-deficient cells was blunted (Fig. 6, C to F). Furthermore, we observed minimal differences in the size and number of SPOT-like structures between uninfected and infected *SAM50*-depleted cells. Mitochondria-*Toxoplasma* contact sites

were unaffected by the loss of *SAM50* (Fig. 6, C to F, and fig. S13). Similar results were obtained in cells depleted of MIC60 (Fig. 6, C to F, and fig. S13). To further test the possibility that SPOTs result from the separation of the OMM and IMM, we generated cells that stably express BFP fused to a tether that bridges the OMM and IMM (O-Iₜ) (41). Expression of the O-Iₜ significantly decreased the rate of formation of SPOTs. In addition, the average number of SPOTs per cell and their diameter were decreased relative to cells that express matrix-targeted BFP during infection (Fig. 6, I to J). Thus, *Toxoplasma* induces SPOTs in a SAM50-dependent manner, and the loss of SAM50 is sufficient to induce the formation of SPOT-like structures in the absence of infection.

Import-linked OMM stress induces OMM shedding independently of infection

If the remodeling of the OMM that occurs during infection represents a general response to import stress, we expected to see the formation of SPOT-like structures upon perturbation of import independently of infection, such as by the overexpression of an OMM protein. To this end, we isolated cells that express the GFP-tagged OMM-targeting α-helical TM of OMP25 at >20-fold times; greater (OMMʰⁱ) than that in cells used in prior experiments (OMMˡᵒ) and that slowed cell proliferation (fig. S15, A and B). We observed a substantial remodeling of the OMM in >20% of OMMʰⁱ cells but not in OMMˡᵒ cells, despite the overexpression of matrix-BFP in both cell lines (fig. S15, C to F). This indicated a link between stress related to the import of OMM proteins and the formation of SPOT-like structures. No remodeling of the OMM was observed after treatment with the protonophore CCCP (carbonyl cyanide m-chlorophenyl hydrazone), which dissipates mitochondrial membrane potential (MMP) and prevents the import of presequence-containing proteins, or MitobloCK6, which inhibits the IMS import pathway (fig. S16) (42). We next generated cells in which we expressed GFP fused to an artificial clogger of the TOM40 channel that forms the entry gate for almost all mitochondrial proteins other than α-helical OMM proteins (fig. S17A). The clogger contains DHFR between a cytochrome b_2 presequence that directs it to the IMM by way of the TOM complex and a heme binding domain that slows its import (fig. S17A) (43). The addition of methotrexate (MTX) that stabilizes the folding state of DHFR leads to its arrest at the TOM complex and further accumulation in the cytosol, as we observed (fig. S17) (43, 44). However, the expression of the clogger with vehicle or MTX did not lead to SPOT formation (fig. S17, C to F). Thus, the OMM is remodeled after stress linked to the overexpression of an α-helical OMM protein
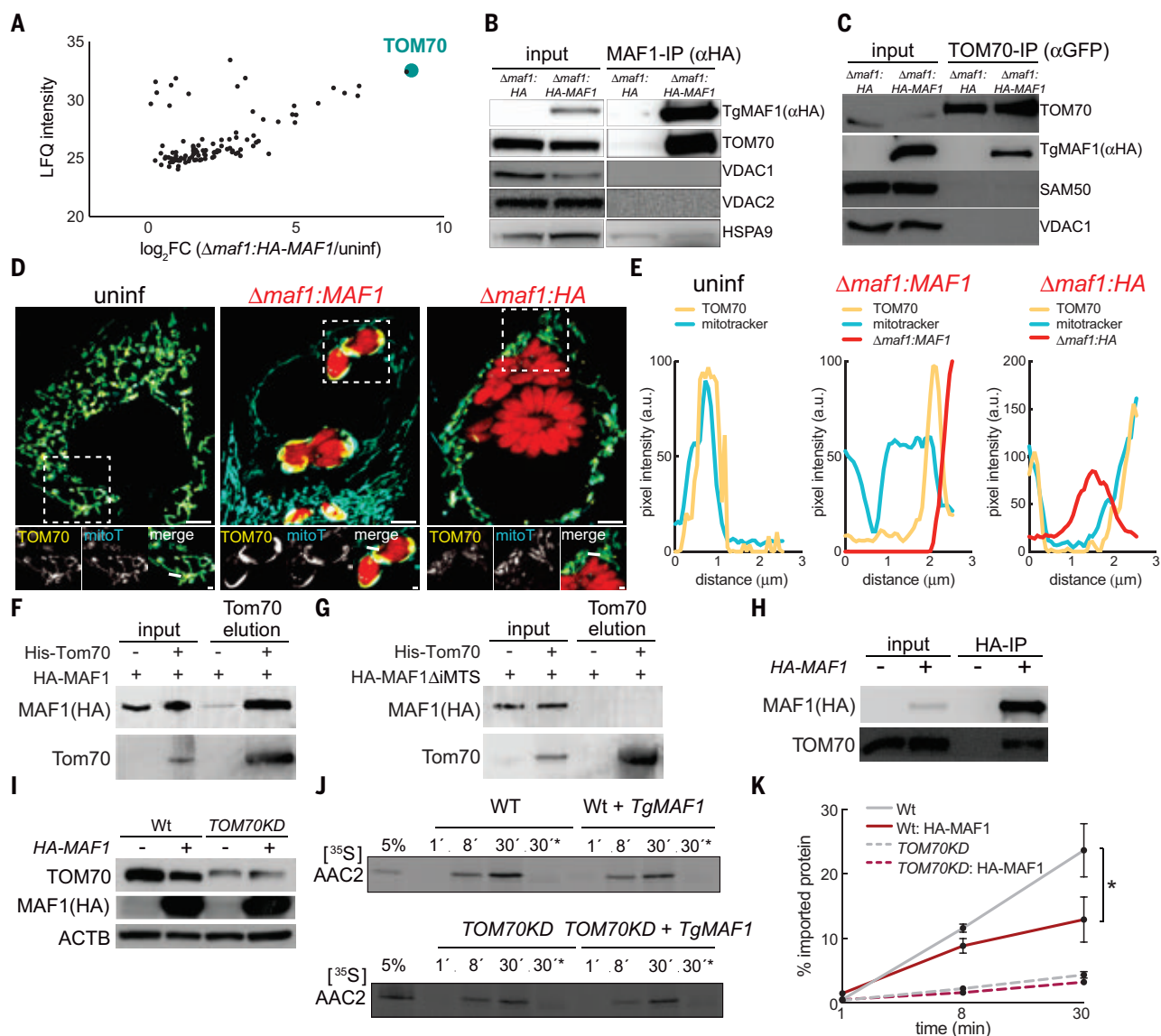
Fig. 4. TgMAF1 binds the host receptor TOM70 and inhibits its import function. (**A**) Anti-HA immunoprecipitates (IPs) from cells that were mock-infected (uninf) or infected with Δ*maf1*:HA-MAF1 parasites at an MOI of 1, 2.4, and 6 and analyzed by means of mass spectrometry; data for MOI 2.4/uninf are shown for 101 human protein hits that had a positive log$_2$FC for the comparisons: all MOIs/uninf, MOI:6/MOI:2.4, MOI:6/MOI:1, MOI:2.4/MOI:1. LFQ, label-free quantification. (**B**) Anti-HA IPs from U2OS cells infected with Δ*maf1*:HA or Δ*maf1*:HA-MAF1 parasites and analyzed by means of immunoblotting (IB) for TgMAF1, ~60 kDa; HSPA9, ~90 kDa; VDAC1, ~34 kda; VDAC2, ~34 kda; TOM70, ~72 kDa. (**C**) Anti-GFP IPs from TOM70-GFP–expressing MEFs 24 hours after infection with Δ*maf1*:HA or Δ*maf1*:HA-MAF1 and analyzed by means of IB for indicated proteins: TOM70-GFP, ~105 kDa; TgMAF1, ~60 kda; SAM50, ~55 kda; VDAC1, ~34 kda. (**D**) Representative live-cell images of mitoT-labeled TOM70-GFP MEFs at 24 hours after mock infection (uninf) or infection with Δ*maf1*:HA-MAF1 or Δ*maf1*:HA parasites.

Scale bars, 5 μm and (inset) 1 μm. (**E**) Corresponding pixel intensity plots for white line in (D) inset. (**F**) HA-MAF1 and (**G**) HA-MAFΔiMTS were incubated with his-tagged TOM70 (cytosolic domain) and subjected to affinity purification with Ni-NTA agarose. Input and elution were analyzed by means of IB. (**H**) HA-IPs from MEFs expressing HA-MAF1 and analyzed by means of IB for TOM70, ~72 kDa, and HA-MAF1, ~60 kDa. (**I**) IB analyses of lysates from WT and TOM70-suppressed 293Ts ± HA-MAF1 cDNA: TOM70, ~72 kDa; HA-MAF1, ~60 kDa. (**J**) [^{35}S]AAC2 import into mitochondria isolated from cells in (I) at indicated times was analyzed by means of SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography. Nonimported proteins were removed through proteinase K treatment; asterisk indicates CCCP treatment. (**K**) Signals in (J) were quantified, and the amount of imported protein relative to input (5%) was plotted. Data are mean ± SEM from three biological replicates, *$P < 0.05$ for WT versus WT:HA-MAF1 by means of two-way ANOVA analysis.

but not depolarization of the MMP, inhibition of the IMS pathway, or clogging of the TOM40 channel.

Whether mitochondria are physically remodeled during physiologically occurring stress at the OMM has been unclear. Here, we describe the discovery of SPOTs, a mechanism of OMM remodeling during *Toxoplasma* infection. We propose that *Toxoplasma* coopts the host TOM70 receptor and SAM50 translocase to promote its insertion into the OMM (Fig. 6K). This enables *Toxoplasma* to hijack a host response to OMM stress (Fig. 6K). During import stress independent of infection, the formation of SPOT-like structures could safe-guard OMM function by sequestering defective import machinery, preventing the accumulation of precursors and misfolded proteins at
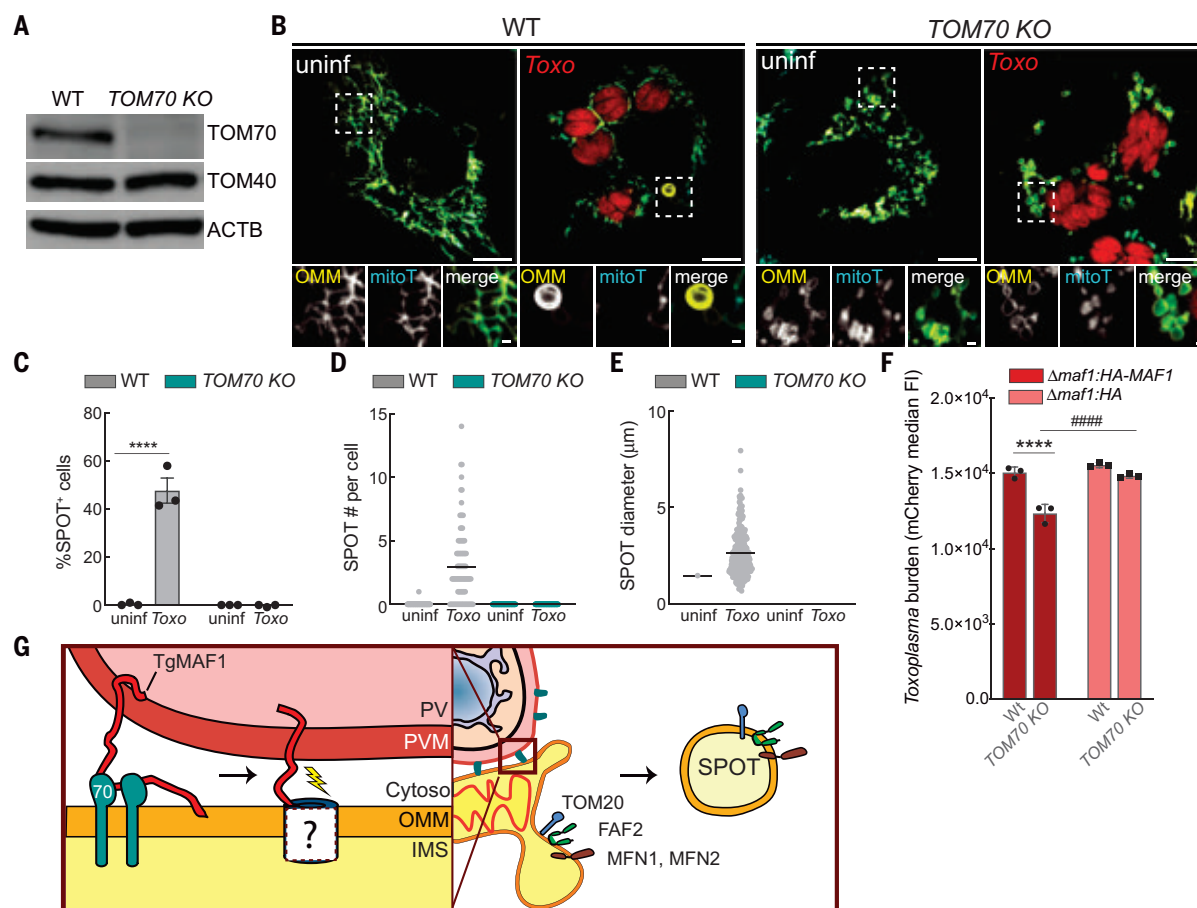
Fig. 5. TOM70 is required for SPOT formation and MAF1-dependent growth.
(**A**) WT and *TOM70*-deleted (*TOM70 KO*) HeLas were analyzed by means of immunoblotting for TOM70, ~72 kDa; TOM40, ~40 kDa; and ACTB, ~45 kDa. (**B**) Representative live-cell images of the OMM (GFP) in uninfected (uninf) and *Toxoplasma* (mCh)–infected (*Toxo*)WT and *TOM70 KO* HeLas labeled with mitoT. Scale bars, 5 μm and (inset) 1 μm. (**C**) Percentage of SPOT-positive cells in experiments as in (B). Data are mean ± SEM of more than 100 cells counted from three biological replicates; ****$P < 0.0001$ for uninfected versus infected by means of two-way ANOVA analysis. (**D** and **E**) Scatterplots with mean (D) number and (E) diameter of SPOTs in experiments as in (B) from more than 30 infected cells from three biological replicates. (**F**) WT and *TOM70 KO* HeLas

were infected with *Δmaf1:HA-MAF1* or *Δmaf1:HA* parasites, rinsed at 2 hours after infection, and analyzed 24 hours after infection by means of flow cytometry for *Toxoplasma* burden [red fluorescent protein (RFP) median FI]. Data are mean ± SEM of three biological experiments, ****$P < 0.0001$ for WT versus *TOM70 KO*, ####$P < 0.0001$ for *Δmaf1:HA_MAF1* versus *Δmaf1:HA* by means of two-way ANOVA analysis. (**G**) Cartoon model of SPOT formation. Host TOM70 mediates insertion of TgMAF1 into host mitochondria through a hypothetical OMM translocase-insertase inducing a stress that leads to SPOT formation and sequestration of import machinery (TOM20) as well as proteins such as FAF2, MFN1, and MFN2 on SPOTs. PV, parasite vacuole; PVM, PV membrane; IMS, intermembrane space.

the OMM that can have toxic consequences for the cell (fig. S18) (*4*, *37*). In the context of infection, however, the constant formation of SPOTs mediates the removal of OMM proteins, including SAM50 and TOM20, which are key components of import machinery, and MFN1 and MFN2, which mediate mitochondrial nutrient competition (Fig. 6K) (*5*).

Discussion

Our data showing that stress linked to the import of OMM proteins leads to the shedding of the OMM during and independently of infection opens several questions, including whether differences in the underlying mechanism of OMM remodeling in these scenarios exist. OMM-GFP expression induced the formation of structures in a manner consistent with our

description of infection-induced SPOTs. However, the OMM was remodeled in a different manner during *SAM50* silencing/depletion. Whether the smaller *SAM50* silencing/depletion–induced structures are a consequence of defective import of OMM proteins that mediate the emergence of SPOTs or represent other classes of structures that emerge from the OMM, such as MDCs and MDVs, is unclear. Furthermore, what is the nature of the TgMAF1- or OMM-GFP–derived stress that leads to OMM remodeling? Although most known for its role in the biogenesis of β-barrel proteins, SAM50 in yeast can cooperate with the mitochondrial import (MIM) complex that inserts C-tail α-helical proteins into the OMM (*45*). Thus, similar machinery in mammals—so far unknown—may facilitate an interaction between the C-terminal

α-helix of TgMAF1 or the OMM-targeting α-helical TM domain of OMM-GFP with SAM50 (*28*). Because SAM50 is the only component of the OMM import machinery with a defined role in bridging the OMM and IMM, perhaps SAM50 can thus function as a sensor that translates stress linked to the import of OMM proteins into a removal of compromised import machinery of the OMM through the disruption of the MIB complex. Additionally, ER whorls that are morphologically similar to SPOTs and contain the ER import translocon have been observed after induction of ER stress (*46*).

What regulates the extent to which the OMM is shed? SNX9 is required for the emergence of a subclass of MDVs that are uniformly shaped and range between 80 and 120 nm in diameter (*22–24*). We found that infection induced much

Fig. 6. SAM50 loss of function mediates SPOT formation during infection. (**A**) Anti-HA IPs were prepared from WT and *TOM70 KO* HeLas infected with Δ*maf1*:HA or Δ*maf1*:HA-MAF1 parasites and analyzed for TgMAF1, ~60 kDa; TOM70, ~72 kDa; SAM50, ~55 kDa; MIC60, ~88 kDa; MIC19, ~25 kDa; TOM40, ~40 kDa. (**B**) Representative IF images of uninf and *Toxo* HFFs at 24 hours after infection. (Inset) SPOT in *Toxo*-infected cell contains SAM50 but not MIC60 or ATP51B (IMM). Scale bars, 5 μm and (inset) 1 μm. (**C**) Representative live-cell images of the OMM (BFP) in uninf and *Toxo*-infected *CTRL*, *TOM70*-, *SAM50*-, and *MIC60*-suppressed (KD) HeLas labeled with mitoT. (Insets) SPOTs are indicated with arrowheads. Scale bars, 5 μm and (inset) 1 μm. (**D**) Percentage (%) of SPOT-positive cells in experiments as in (C); data are mean ± SEM of more than 100 cells counted from *n* = 3 replicates; *P < 0.05, **P < 0.01, ****P < 0.0001 for uninf versus inf, ####P < 0.0001 for *CTRL KD* versus *SAM50 KD*, *MIC60 KD* by means of two-way ANOVA analysis. (**E** and **F**) Scatterplots with mean (E) number and (F) diameter of SPOTs in experiments as in (C) from more than 30 infected cells from three replicates. (**G**) Representative live-cell images of the OMM (GFP) in uninf and *Toxoplasma* (mCh)–infected U2OS cells expressing matrix-BFP (matrix) or BFP fused to an OMM-IMM tether (O-l_t). (**H**) Percent of SPOT-positive cells in experiments as in (G). Data are mean ± SEM of more than 100 cells counted from *n* = 3 biological replicates; ***P < 0.001, ****P < 0.0001 for uninf versus inf, ####P < 0.0001 for matrix versus O-l_t by two-way ANOVA analysis. (**I** and **J**) Scatterplots with mean (I) number and (J) diameter

of SPOTs in experiments as in (G) from more than 30 infected cells from three replicates. (**K**) Model of SPOT formation: TOM70 (70) mediates the interaction between MAF1 and SAM50 and/or a hypothetical translocase (?), which induces a disassembly of the MIB complex (SAM50, MIC60, and MIC19) and SPOT formation. The constitutive shedding of SPOTs depletes OMM proteins that restrict parasite growth (MFN1 and MFN2) and sequesters import machinery required for mitochondrial biogenesis on SPOTs.

larger SPOTs that range up to 10 μm in diameter, are also SNX9-dependent, and vary in substructure and shape. What factors dictate their morphology and the fate of their proteins? Although MFN1 and MFN2 are targeted for degradation, it remains possible that other SPOT-localized proteins such as SAM50, TOM70, and TOM20 are recycled into the mitochondrial network. SPOTs being oftentimes multivesicular and containing proteins required for fusion (MFNs) and trafficking (MIROs) raises the possibility that these structures might be able to fuse with one another, or with mitochondria to reintegrate into the mitochondrial network.
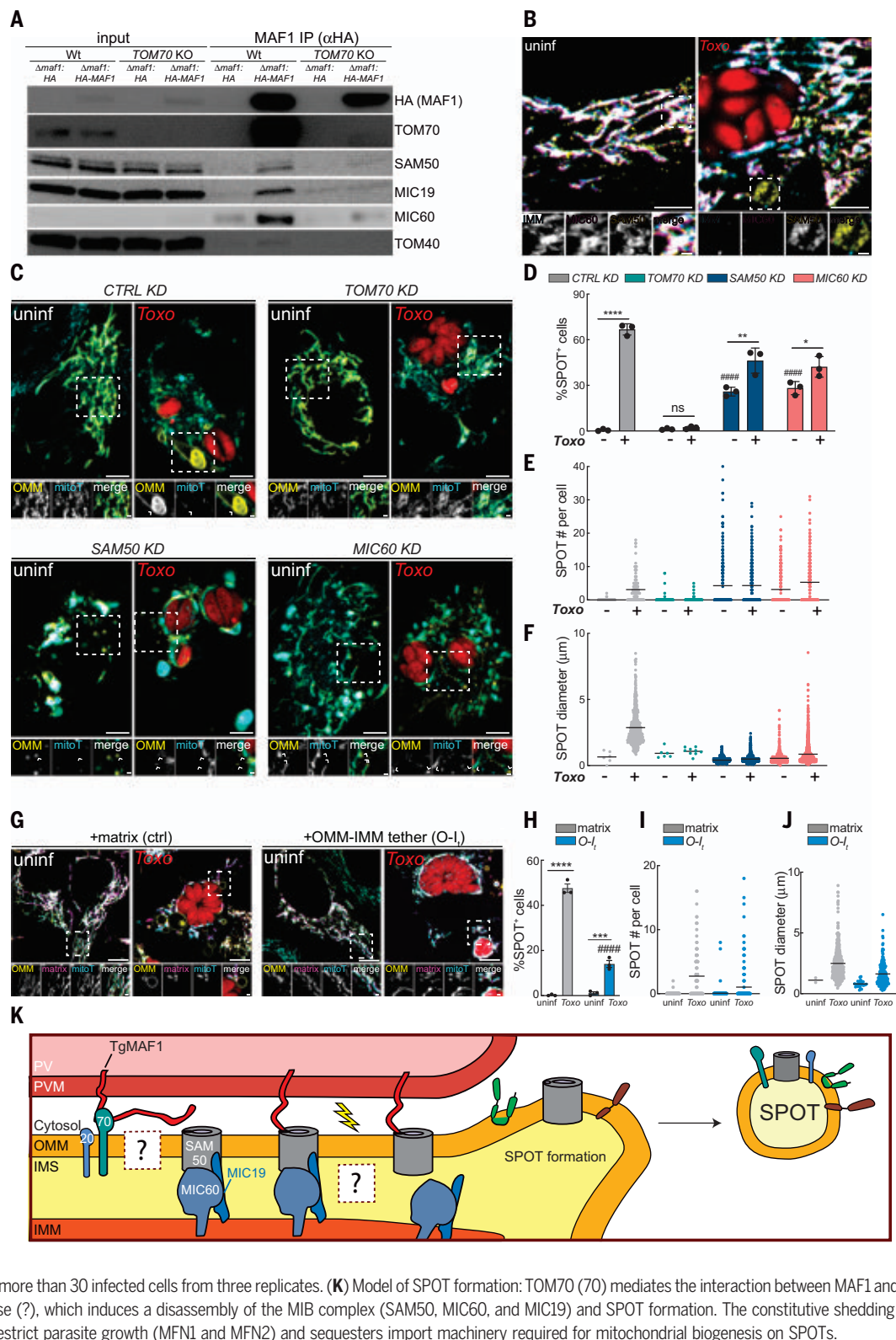
Whether other pathogen effector proteins coopt host import receptors for their function, analogous to how TgMAF1 exploits host TOM70, is little explored. Pathogens such as *Toxoplasma* might target SAM50 and TOM70, key regulators of mitochondrial import, to impede the biogenesis of organelles that function as nutrient competitors or immune signaling hubs during infection (*5–7*). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) encodes for a protein of unknown function (Orf9b) that binds TOM70 to suppress antiviral interferon responses (*47*). Host TOM complex components also mediate contact sites between mitochondria and the vacuole in which *Chlamydia* resides, but whether this bacterium inhibits TOM receptors or induces SPOT formation is not known (*48*).

Our study of the interaction between the human parasite *Toxoplasma* and host mitochondria led to the discovery of a mechanism by which the OMM is remodeled: the formation of SPOTs. These findings shed light on a potentially broader mechanism of organellar response to import-related stress and reveal a strategy by which diverse pathogens may disrupt mitochondrial function.

Materials and methods
Cell culture and cell lines

All cell lines were cultured in complete DMEM (cDMEM: DMEM and 10% heat-inactivated FBS). Cells were tested every 2 weeks for *Mycoplasma* infection by means of polymerase chain reaction (PCR). Details of cell lines generated (including single-guide RNA sequences), used, and origin are in the supplementary materials.

Parasite culture and strains

Toxoplasma gondii parasites were maintained by serial passage in human foreskin fibroblast (HFF) monolayers in cDMEM. Details of parasite lines generated, used, and origin are in the supplementary materials.

Plasmid descriptions, transfections, and siRNA treatment

For transient expression, cells were transfected ~12 hours prior to infection using X-tremeGene reagent (Sigma) per manufacturer's instructions, lipofectamine RNAiMax (Invitrogen) was used per manufacturer's instructions for siRNA treatment. Details of plasmids generated, used, and origin are in the supplementary materials.

Lentiviral production generation of cells stably expressing cDNAs

For production of 293T human embryonic kidney (HEK) cells were transfected using the X-tremeGENE 9 DNA Transfection Reagent (Roche) with 1 µg psPAX2 packaging vector (Addgene #12260), 0.3 µg pCMV-VSVG envelope vector (Addgene #8454) and 1 µg of the relevant plasmid of interest. The next day, the medium was exchanged, and two days post transfection, the virus-containing supernatant was filtered with a 0.45 µM filter and supplemented with polybrene to a final concentration of 5 µg/ml. The virus-containing filtrate was added to 50,000 target cells and exchanged for cDMEM the next day. MEFs, HeLas, U2OS cells were subsequently selected with 8 to 10 µg/ml blasticidin, or 1-2 µg/ml puromycin for 3-5 days. To isolate cells positive for expression of fluorescent protein, transduced cells were sorted for low and high GFP positivity as indicated.

Live cell imaging

Cells were plated on 35mm, 6-well, or 24-well CELLview glass bottom cell culture dishes (Greiner Bio-One), treated as indicated in text (i.e., infection, transfection, 25 mM CCCP treatment for 30 min, or fixation and permeabilization while imaging), and imaged using an Olympus IXplore SpinSR 50 mm spinning disk confocal microscope. Live cell imaging was performed in cDMEM with incubation at 37°C and 5% CO_2. All images were taken with a 100X/1.35 silicon oil objective and excitation with either 405, 488, 561, or 640 laser lines, using ORCA-Flash4.0 cameras (Hamatsu), and cellSens Software. Details of mitotracker and FL-HPC labeling, immunofluorescence analysis, and antibodies used are in the supplementary materials.

Yeast strains, growth conditions, and expression plasmids

The yeast strains used in this study are based on *Saccharomyces cerevisiae* BY4741 (EUROSCARF). Yeast cells were cultured using standard protocols on SC-Leucin (0.67% [w/v] yeast nitrogen base with ammonium sulfate; 0.07% [w/v] amino acid mixture) with 2% [w/v] galactose. Cultures were grown at 30°C until early logarithmic growth phase, which was determined based on the optical density at a wavelength of 600 nm. Full details of strains used and generated, isolation of yeast and mammalian mitochondria, affinity purification of HA-MAF1 from yeast, in vitro binding to TOM70, and in vitro protein import into yeast and mammalian mitochondria in supplementary material.

Line scan analyses

Line-scan analysis of relative fluorescence intensity was performed by measuring pixel intensity across an indicated line using Fiji software.

Immunoprecipitation of mitochondria

Seven million to 8 million MEFs or 293Ts cells were plated in a 15 cm dish and the following day, cells were either mock-infected or infected with *Toxoplasma* (RH*Δku80:mCherry*⁺) at a multiplicity of infection (MOI) of 5. 24 hpi, cells were scraped rinsed 2X in chilled 1X PBS, scraped in 1X chilled PBS supplemented with a protease and inhibitor cocktail (Thermo Scientific a32961) and Phosstop phosphatase inhibitors (Sigma 4906837001)(1XPBS+Inh). Harvested cells were centrifuged at 1000xg for 2 min at 4C and resuspended in 1 ml 1X PBS with inhibitors. 50 µl were collected for the whole cell fraction and the rest was triturated 10X on ice using a 1 ml syringe and $27^{3/4}$ gauge needle. Lysed cells were centrifuged at 1000xg for 2 min at 4C. Cleared lysate was added to 150ul anti-HA bead slurry that had previously been rinsed 3X in 1X PBS and resuspended in 100 µl 1X PBS. Subsequently, mitochondria were immunopurified as previously described and processed for immunoblot analysis or mass spectrometry analysis (*13*). More details on immunoblot analysis and proteomics sample preparation are provided in the supplementary materials.

Flow cytometry analysis

Monolayers of U2OS cells expressing OMM-GFP were rinsed with PBS, trypsinized and fixed in 2% paraformaldehyde in FACS buffer (3% FBS in PBS) for 10 min. After a brief spin, cells were resuspended in FACS buffer and sorted on a FACSAria Fusion Cell Sorter (BD Biosciences) GFP mean fluorescence intensity (mFI) using BD FACSDiva software. ~250,000 cells were sorted into low and high (20× greater than low) GFP bins and transferred to a cell culture dish for continued passaging.

Immunoprecipitation of HA-MAF1 or TOM70

Seven million to 8 million TOM70-expressing MEFs or U2OS cells were infected with *Toxoplasma* at an MOI of 7. 24 hpi cells were rinsed 2X in chilled 1X PBS, scraped down in chilled 1XPBS+Inh, centrifuged at 1000xg for 2 min, and resuspended in lysis buffer for 15 min at 4°C. Cleared lysates were incubated with either 80 µl magnetic anti-HA-beads (Thermo Scientific) or 25 µl magnetic anti-GFP-nanobodies (Chromotek) overnight. The beads were washed 3x times with 1XPBS+Inh. Afterwards, the samples were eluted from the magnetic beads with 2X SDS buffer by incubating them at 40°C for 10 min. Samples were processed for gel electrophoresis and probed with indicated antibodies. For more details on proteomics sample preparation, please see supplementary materials.

Statistical analyses

All statistical analyses were performed using one-way analysis of variance (ANOVA), two-way ANOVA, or an unpaired *t* test in GraphPad Prism 9 software and are indicated accordingly. Volcano plot rendering of proteomics of whole-cell and immunopurified mitochondria fractions are provided in the main and supplementary figures.

REFERENCES AND NOTES

1. L. Pernas, L. Scorrano, Mito-morphosis: Mitochondrial fusion, fission, and cristae remodeling as key mediators of cellular function. *Annu. Rev. Physiol.* **78**, 505–531 (2016). doi: 10.1146/annurev-physiol-021115-105011; pmid: 26667075

2. S. Rath et al., MitoCarta3.0: An updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res.* **49** (D1), D1541–D1547 (2021). doi: 10.1093/nar/gkaa1011; pmid: 33174596

3. J. A. MacKenzie, R. M. Payne, Mitochondrial protein import and human health and disease. *Biochim. Biophys. Acta* **1772**, 509–523 (2007). doi: 10.1016/j.bbadis.2006.12.002; pmid: 17300922

4. J. Song, J. M. Herrmann, T. Becker, Quality control of the mitochondrial proteome. *Nat. Rev. Mol. Cell Biol.* **22**, 54–70 (2021). doi: 10.1038/s41580-020-00300-2; pmid: 33093673

5. L. Pernas, C. Bean, J. C. Boothroyd, L. Scorrano, Mitochondria restrict growth of the intracellular parasite *Toxoplasma gondii* by limiting its uptake of fatty acids. *Cell Metab.* **27**, 886–897.e4 (2018). doi: 10.1016/j.cmet.2018.02.018; pmid: 29617646

6. V. Tiku, M. W. Tan, I. Dikic, Mitochondrial functions in infection and immunity. *Trends Cell Biol.* **30**, 263–275 (2020). doi: 10.1016/j.tcb.2020.01.006; pmid: 32200805

7. L. Pernas, Cellular metabolism in the defense against microbes. *J. Cell Sci.* **134**, jcs252023 (2021). doi: 10.1242/jcs.252023; pmid: 33558420

8. R. B. Seth, L. Sun, Z. J. Chen, Antiviral innate immunity pathways. *Cell Res.* **16**, 141–147 (2006). doi: 10.1038/sj.cr.7310019; pmid: 16474426

9. T. Rudel, O. Kepp, V. Kozjak-Pavlovic, Interactions between bacterial pathogens and mitochondrial cell death pathways. *Nat. Rev. Microbiol.* **8**, 693–705 (2010). doi: 10.1038/nrmicro2421; pmid: 20818415

10. J. G. Montoya, O. Liesenfeld, Toxoplasmosis. *Lancet* **363**, 1965–1976 (2004). doi: 10.1016/S0140-6736(04)16412-X; pmid: 15194258

11. T. C. Medeiros, C. Mehra, L. Pernas, Contact and competition between mitochondria and microbes. *Curr. Opin. Microbiol.* **63**, 189–194 (2021). doi: 10.1016/j.mib.2021.07.014; pmid: 34411806

12. M. Dumoux, R. D. Hayward, Membrane contact sites between pathogen-containing compartments and host organelles. *Biochim. Biophys. Acta* **1861** (8 Pt B), 895–899 (2016). doi: 10.1016/j.bbalip.2016.01.018; pmid: 26825687

13. W. W. Chen, E. Freinkman, D. M. Sabatini, Rapid immunopurification of mitochondria for metabolite profiling and absolute quantification of matrix metabolites. *Nat. Protoc.* **12**, 2215–2231 (2017). doi: 10.1038/nprot.2017.104; pmid: 29532801

14. V. Soubannier et al., A vesicular transport pathway shuttles cargo from mitochondria to lysosomes. *Curr. Biol.* **22**, 135–141 (2012). doi: 10.1016/j.cub.2011.11.057; pmid: 22226745

15. A. L. Hughes, C. E. Hughes, K. A. Henderson, N. Yazvenko, D. E. Gottschling, Selective sorting and destruction of mitochondrial membrane proteins in aged yeast. *eLife* **5**, e13943 (2016). doi: 10.7554/eLife.13943; pmid: 27097106

16. A. M. English et al., ER-mitochondria contacts promote mitochondrial-derived compartment biogenesis. *J. Cell Biol.* **219**, e202002144 (2020). doi: 10.1083/jcb.202002144; pmid: 33090183

17. M. H. Schuler et al., Mitochondrial-derived compartments facilitate cellular adaptation to amino acid stress. *Mol. Cell* **81**, 3786–3802.e13 (2021). doi: 10.1016/j.molcel.2021.08.021; pmid: 34547239

18. M. Neuspiel et al., Cargo-selected transport from the mitochondria to peroxisomes is mediated by vesicular carriers. *Curr. Biol.* **18**, 102–108 (2008). doi: 10.1016/j.cub.2007.12.038; pmid: 18207745

19. M. H. Schuler, A. M. English, L. VanderMeer, J. M. Shaw, A. L. Hughes, Amino acids promote mitochondrial-derived compartment formation in mammalian cells. bioRxiv [Preprint] 23 December 2020. .doi: 10.1101/2020.12.23.424218

20. G. L. McLelland, V. Soubannier, C. X. Chen, H. M. McBride, E. A. Fon, Parkin and PINK1 function in a vesicular trafficking pathway regulating mitochondrial quality control. *EMBO J.* **33**, 282–295 (2014). doi: 10.1002/embj.201385902; pmid: 24446486

21. S. R. Denison et al., Alterations in the common fragile site gene Parkin in ovarian and other cancers. *Oncogene* **22**, 8370–8378 (2003). doi: 10.1038/sj.onc.1207072; pmid: 14614460

22. D. Matheoud et al., Parkinson's disease-related proteins PINK1 and parkin repress mitochondrial antigen presentation. *Cell* **166**, 314–327 (2016). doi: 10.1016/j.cell.2016.05.039; pmid: 27345367

23. R. Lundmark, S. R. Carlsson, SNX9—A prelude to vesicle release. *J. Cell Sci.* **122**, 5–11 (2009). doi: 10.1242/jcs.037135; pmid: 19092055

24. K. Todkar et al., Selective packaging of mitochondrial proteins into extracellular vesicles prevents the release of mitochondrial DAMPs. *Nat. Commun.* **12**, 1971 (2021). doi: 10.1038/s41467-021-21984-w; pmid: 33785738

25. A. P. Sinai, P. Webster, K. A. Joiner, Association of host cell endoplasmic reticulum and mitochondria with the *Toxoplasma gondii* parasitophorous vacuole membrane: A high affinity interaction. *J. Cell Sci.* **110**, 2117–2128 (1997). doi: 10.1242/jcs.110.17.2117; pmid: 9378762

26. L. Scorrano et al., Coming together to define membrane contact sites. *Nat. Commun.* **10**, 1287 (2019). doi: 10.1038/s41467-019-09253-3; pmid: 30894536

27. L. Pernas et al., Toxoplasma effector MAF1 mediates recruitment of host mitochondria and impacts the host response. *PLOS Biol.* **12**, e1001845 (2014). doi: 10.1371/journal.pbio.1001845; pmid: 24781109

28. M. L. Blank et al., A *Toxoplasma gondii* locus required for the direct manipulation of host mitochondria has maintained multiple ancestral functions. *Mol. Microbiol.* **108**, 519–535 (2018). doi: 10.1111/mmi.13947; pmid: 29505111

29. F. D. Kelly et al., *Toxoplasma gondii* MAF1b binds the host cell MIB complex to mediate mitochondrial association. *MSphere* **2**, e00183-17 (2017). doi: 10.1128/mSphere.00183-17; pmid: 28567444

30. M. Suzuki, O. Danilchanka, J. J. Mekalanos, Vibrio cholerae T3SS effector VopE modulates mitochondrial dynamics and innate immune signaling by targeting Miro GTPases. *Cell Host Microbe* **16**, 581–591 (2014). doi: 10.1016/j.chom.2014.09.015; pmid: 25450857

31. B. Mueller, E. J. Klemm, E. Spooner, J. H. Claessen, H. L. Ploegh, SEL1L nucleates a protein complex required for dislocation of misfolded glycoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12325–12330 (2008). doi: 10.1073/pnas.0805371105; pmid: 18711132

32. N. C. Chan et al., Broad activation of the ubiquitin-proteasome system by Parkin is critical for mitophagy. *Hum. Mol. Genet.* **20**, 1726–1737 (2011). doi: 10.1093/hmg/ddr048; pmid: 21296869

33. S. Nahar, A. Chowdhury, T. Ogura, M. Esaki, A AAA ATPase Cdc48 with a cofactor Ubx2 facilitates ubiquitylation of a mitochondrial fusion-promoting factor Fzo1 for proteasomal degradation. *J. Biochem.* **167**, 279–286 (2020). doi: 10.1093/jb/mvz104; pmid: 31804690

34. M. K. Shaw, C. Y. He, D. S. Roos, L. G. Tilney, Proteasome inhibitors block intracellular growth and replication of *Toxoplasma gondii*. *Parasitology* **121**, 35–47 (2000). doi: 10.1017/S0031182099006071; pmid: 11085223

35. M. L. Blank et al., *Toxoplasma gondii* association with host mitochondria requires key mitochondrial protein import machinery. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2013336118 (2021). doi: 10.1073/pnas.2013336118; pmid: 33723040

36. S. Backes et al., Tom70 enhances mitochondrial preprotein import efficiency by binding to internal targeting sequences. *J. Cell Biol.* **217**, 1369–1382 (2018). doi: 10.1083/jcb.201708044; pmid: 29382700

37. S. Backes et al., The chaperone-binding activity of the mitochondrial surface receptor Tom70 protects the cytosol against mitoprotein-induced stress. *Cell Rep.* **35**, 108936 (2021). doi: 10.1016/j.celrep.2021.108936; pmid: 33826901

38. A. C. Fan, M. K. Bhangoo, J. C. Young, Hsp90 functions in the targeting and outer membrane translocation steps of Tom70-mediated mitochondrial import. *J. Biol. Chem.* **281**, 33313–33324 (2006). doi: 10.1074/jbc.M605250200; pmid: 16968702

39. H. Yamamoto et al., Roles of Tom70 in import of presequence-containing mitochondrial proteins. *J. Biol. Chem.* **284**, 31635–31646 (2009). doi: 10.1074/jbc.M109.041756; pmid: 19767391

40. C. Ott et al., Sam50 functions in mitochondrial intermembrane space bridging and biogenesis of respiratory complexes. *Mol. Cell. Biol.* **32**, 1173–1188 (2012). doi: 10.1128/MCB.06388-11; pmid: 22252321

41. M. P. Viana, R. M. Levytskyy, R. Anand, A. S. Reichert, O. Khalimonchuk, Protease OMA1 modulates mitochondrial bioenergetics and ultrastructure through dynamic association with MICOS complex. *iScience* **24**, 102119 (2021). doi: 10.1016/j.isci.2021.102119; pmid: 33644718

42. D. V. Dabir et al., A small molecule inhibitor of redox-regulated protein translocation into mitochondria. *Dev. Cell* **25**, 81–92 (2013). doi: 10.1016/j.devcel.2013.03.006; pmid: 23597483

43. U. Bömer et al., The sorting route of cytochrome b2 branches from the general mitochondrial import pathway at the preprotein translocase of the inner membrane. *J. Biol. Chem.* **272**, 30439–30446 (1997). doi: 10.1074/jbc.272.48.30439; pmid: 9374535

44. M. Eilers, G. Schatz, Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria. *Nature* **322**, 228–232 (1986). doi: 10.1038/322228a0; pmid: 3016548

45. K. N. Doan et al., The mitochondrial import complex MIM functions as main translocase for α-helical outer membrane proteins. *Cell Rep.* **31**, 107567 (2020). doi: 10.1016/j.celrep.2020.107567; pmid: 32348752

46. F. Xu et al., COPII mitigates ER stress by promoting formation of ER whorls. *Cell Res.* **31**, 141–156 (2021). doi: 10.1038/s41422-020-00416-2; pmid: 32989223

47. C. U. Mårtensson et al., Mitochondrial protein translocation-associated degradation. *Nature* **569**, 679–683 (2019). doi: 10.1038/s41586-019-1227-y; pmid: 31118508

48. I. Derré, M. Pypaert, A. Dautry-Varsat, H. Agaisse, RNAi screen in Drosophila cells reveals the involvement of the Tom complex in Chlamydia infection. *PLOS Pathog.* **3**, 1446–1458 (2007). doi: 10.1371/journal.ppat.0030155; pmid: 17967059

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi4343
Materials and Methods
Figs. S1 to S18
References (49–62)
Data Files S1 to S3
Movies S1 and S2

CORONAVIRUS

Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh

Jason Abaluck*†, Laura H. Kwong†, Ashley Styczynski†, Ashraful Haque, Md. Alamgir Kabir, Ellen Bates-Jefferys, Emily Crawford, Jade Benjamin-Chung, Shabib Raihan, Shadman Rahman, Salim Benhachmi, Neeti Zaman Bintee, Peter J. Winch, Maqsud Hossain, Hasan Mahmud Reza, Abdullah All Jaber, Shawkee Gulshan Momen, Aura Rahman, Faika Laz Banti, Tahrima Saiha Huq, Stephen P. Luby‡, Ahmed Mushfiq Mobarak*‡

INTRODUCTION: Mask usage remains low across many parts of the world during the COVID-19 pandemic, and strategies to increase mask-wearing remain untested. Our objectives were to identify strategies that can persistently increase mask-wearing and assess the impact of increasing mask-wearing on symptomatic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections.

RATIONALE: We conducted a cluster-randomized trial of community-level mask promotion in rural Bangladesh from November 2020 to April 2021 (N = 600 villages, N = 342,183 adults). We cross-randomized mask promotion strategies at the village and household level, including cloth versus surgical masks. All intervention arms received free masks, information on the importance of masking, role modeling by community leaders, and in-person reminders for 8 weeks. The control group did not receive any interventions. Participants and surveillance staff were not informed of treatment assignments, but project materials were clearly visible. Outcomes included symptomatic SARS-CoV-2 seroprevalence (primary) and prevalence of proper mask-

wearing, physical distancing, social distancing, and symptoms consistent with COVID-19 illness (secondary). Mask-wearing and distancing were assessed through direct observation at least weekly at mosques, markets, the main entrance roads to villages, and tea stalls. Individuals were coded as physically distanced if they were at least one arm's length from the nearest adult; social distancing was measured using the total number of adults observed in public areas. At 5- and 9-week follow-ups, we surveyed all reachable participants about COVID-19–related symptoms. Blood samples collected at 10- to 12-week follow-ups for symptomatic individuals were analyzed for SARS-CoV-2 immunoglobulin G (IgG) antibodies.

RESULTS: There were 178,322 individuals in the intervention group and 163,861 individuals in the control group. The intervention increased proper mask-wearing from 13.3% in control villages (N = 806,547 observations) to 42.3% in treatment villages (N = 797,715 observations) (adjusted percentage point difference = 0.29; 95% confidence interval = [0.26, 0.31]). This tripling of mask usage was

sustained during the intervention period and for 2 weeks after. Physical distancing increased from 24.1% in control villages to 29.2% in treatment villages (adjusted percentage point difference = 0.05 [0.04, 0.06]). We saw no change in social distancing. After 5 months, the impact of the intervention on mask-wearing waned, but mask-wearing remained 10 percentage points higher in the intervention group. Beyond the core intervention of free distribution and promotion at households, mosques, and markets; leader endorsements; and periodic monitoring and reminders, several elements had no additional effect on mask-wearing, including text reminders, public signage commitments, monetary or nonmonetary incentives, and altruistic messaging or verbal commitments.

The proportion of individuals with COVID-19–like symptoms was 7.63% (N = 12,784) in the intervention arm and 8.60% (N = 13,287) in the control arm, an estimated 11.6% reduction after controlling for baseline covariates. Blood samples were collected from consenting, symptomatic adults (N = 10,790). Adjusting for baseline covariates, the intervention reduced symptomatic seroprevalence by 9.5% (adjusted prevalence ratio = 0.91 [0.82, 1.00]; control prevalence = 0.76%; treatment prevalence = 0.68%). We find that surgical masks are particularly effective in reducing symptomatic seroprevalence of SARS-CoV-2. In villages randomized to surgical masks (N = 200), the relative reduction was 11.1% overall (adjusted prevalence ratio = 0.89 [0.78, 1.00]). The effect of the intervention is most concentrated among the elderly population; in surgical mask villages, we observe a 35.3% reduction in symptomatic seroprevalence among individuals ≥60 years old (adjusted prevalence ratio = 0.65 [0.45, 0.85]). We see larger reductions in symptoms and symptomatic seropositivity in villages that experienced larger increases in mask use. No adverse events were reported.

CONCLUSION: A randomized-trial of community-level mask promotion in rural Bangladesh during the COVID-19 pandemic shows that the intervention increased mask usage and reduced symptomatic SARS-CoV-2 infections, demonstrating that promoting community mask-wearing can improve public health. ∎
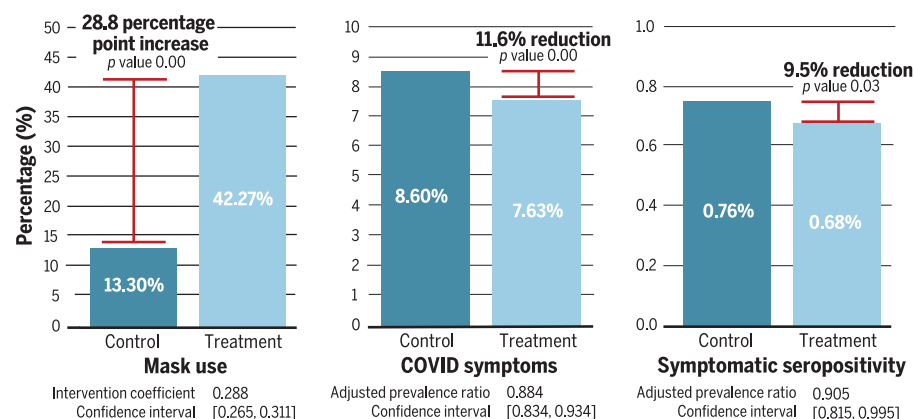
Impact of intervention on mask use and biological outcomes. The figure shows the raw means of mask-wearing (left), COVID-19 symptoms (middle), and symptomatic seropositivity (right) in the control and treatment arms. The estimated change in each outcome, confidence intervals, and p values adjust for preregistered covariates (and thus are not computable from the raw values). Individuals who were symptomatic but did not consent to blood collection were dropped from the sample; measured symptomatic seropositivity thus understates the true fraction of the population that was symptomatic seropositive.

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abi9069

CORONAVIRUS

Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh

Jason Abaluck[1]*†, Laura H. Kwong[2,3]†, Ashley Styczynski[4]†, Ashraful Haque[5], Md. Alamgir Kabir[5], Ellen Bates-Jefferys[6], Emily Crawford[1], Jade Benjamin-Chung[7], Shabib Raihan[5], Shadman Rahman[5], Salim Benhachmi[8], Neeti Zaman Bintee[5], Peter J. Winch[9], Maqsud Hossain[10], Hasan Mahmud Reza[11], Abdullah All Jaber[10], Shawkee Gulshan Momen[10], Aura Rahman[10], Faika Laz Banti[10], Tahrima Saiha Huq[10], Stephen P. Luby[2,4]‡, Ahmed Mushfiq Mobarak[1,12]*‡

We conducted a cluster-randomized trial to measure the effect of community-level mask distribution and promotion on symptomatic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in rural Bangladesh from November 2020 to April 2021 (N = 600 villages, N = 342,183 adults). We cross-randomized mask type (cloth versus surgical) and promotion strategies at the village and household level. Proper mask-wearing increased from 13.3% in the control group to 42.3% in the intervention arm (adjusted percentage point difference = 0.29; 95% confidence interval = [0.26, 0.31]). The intervention reduced symptomatic seroprevalence (adjusted prevalence ratio = 0.91 [0.82, 1.00]), especially among adults ≥60 years old in villages where surgical masks were distributed (adjusted prevalence ratio = 0.65 [0.45, 0.85]). Mask distribution with promotion was a scalable and effective method to reduce symptomatic SARS-CoV-2 infections.

A s of September 2021, the COVID-19 pandemic has taken the lives of more than 4.7 million people. Inspired by the growing body of scientific evidence that face masks have the potential to slow the spread of the disease and save lives (1–10), we conducted a cluster-randomized controlled trial covering 342,183 adults in 600 villages in rural Bangladesh with the dual goals of (i) identifying strategies to increase community-wide mask-wearing and (ii) tracking changes in symptomatic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections as a result of our intervention. Although vaccines may constrain the spread of SARS-CoV-2 in the long-term, it is unlikely that a substantial fraction of the population in low- and middle-income countries will have access to vaccines before the end of 2021 (11). Develop-

ing scalable and effective means of combating COVID-19 is thus of first-order policy importance.

The World Health Organization (WHO) declined to recommend mask adoption until June 2020, citing the lack of evidence from community-based randomized-controlled trials as well as concerns that mask-wearing would create a false sense of security (12). Critics argued that those who wore masks would engage in compensating behaviors, such as failing to physically distance from others, resulting in a net increase in transmission (13). We directly test this hypothesis by measuring physical distancing.

We designed our trial to encourage universal mask-wearing at the community level, rather than mask-wearing among only those with symptoms. We encouraged even healthy individuals to wear masks because a substantial share of COVID-19 transmission stems from asymptomatic or presymptomatic individuals (14) and masks may protect healthy wearers by reducing the inhalation of aerosols or droplets (15–17).

After performing pilot studies, we settled on a core intervention package that combined household mask distribution with communication about the value of mask-wearing; mask promotion and in-person reminders at mosques, markets, and other public places; and role-modeling by public officials and community leaders. We also tested several other strategies in subsamples, such as asking people to make a verbal commitment, creating opportunities for social signaling, text messaging, and providing village-level incentives to increase

mask-wearing. The selection of strategies to test was informed by both our pilot study results and research in public health, psychology (18–20), economics (21–23), marketing (24–26), and other social sciences (27) on product promotion and dissemination strategies. We tested many different strategies because it was difficult to predict in advance which ones would lead to persistent increases in mask-wearing. Prediction studies we conducted with policymakers and public health experts at the WHO, India's National Council of Applied Economic Research, and the World Bank suggested that even these experts with influence over policy design could not easily predict which specific strategies would prove most effective in our trial.

We powered our intervention around the primary outcome of symptomatic seroprevalence. During our study, we collected survey data on the prevalence of WHO-defined COVID-19 symptoms from all available study participants and then collected blood samples at endline from those who reported symptoms at any time during the 8-week study. Our trial is therefore designed to track the fraction of individuals who are both symptomatic and seropositive. We chose this as our primary outcome because (i) the goal of public health policy is ultimately to prevent symptomatic infections (even if preventing asymptomatic infections is instrumentally important in achieving that goal) and (ii) symptomatic individuals are far more likely to be seropositive so powering for this outcome required conducting an order of magnitude fewer costly blood tests. As secondary outcomes, we also report the effects of our intervention on WHO-defined symptoms for probable COVID-19 infection and mask-wearing.

Bangladesh is a densely populated country with 165 million inhabitants; reported infections reached 15,000 per day during our study period, but reported cases and deaths are likely underestimated by one to two orders of magnitude (28–32). The evolution of mask use over time in Bangladesh is discussed in greater detail in (33). In Bangladesh, the government strongly recommended mask use from early April 2020. In an April 2020 telephone survey, more than 80% of respondents self-reported wearing a mask and 97% self-reported owning a mask. The Bangladeshi government formally mandated mask use in late May 2020 and threatened to fine those who did not comply, although enforcement was weak to nonexistent, especially in rural areas. During in-person surveillance between 21 and 25 May 2020 in 1441 places in 52 districts, we observed 51% of about 152,000 individuals wearing a mask. Another wave of surveillance was conducted between 19 and 22 June 2020 in the same 1441 locations, and mask-wearing dropped to 26%, with 20% wearing masks that covered their mouth and

[1]Yale School of Management, Yale University, New Haven, CT, USA. [2]Woods Institute for the Environment, Stanford University, Stanford, CA, USA. [3]Division of Environmental Health Sciences, University of California, Berkeley, Berkeley, CA, USA. [4]Division of Infectious Diseases and Geographic Medicine, Stanford University, Stanford, CA, USA. [5]Innovations for Poverty Action Bangladesh, Dhaka, Bangladesh. [6]Innovations for Poverty Action, Evanston, IL, USA. [7]Department of Epidemiology and Population Health, School of Medicine, Stanford University, Stanford, CA, USA. [8]Yale Research Initiative on Innovation and Scale, Yale University, New Haven, CT, USA. [9]Social and Behavioral Interventions Program, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [10]NGRI, North South University, Dhaka, Bangladesh. [11]Department of Pharmaceutical Sciences, North South University, Dhaka, Bangladesh. [12]Department of Economics, Deakin University, Melbourne, Australia.
*Corresponding author. Email: jason.abaluck@yale.edu (J.A.); ahmed.mobarak@yale.edu (A.M.M.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

nose and 6% wearing masks improperly. An August 2020 phone survey in rural Kenya found that although 88% of respondents claim to wear masks in public, direct observation revealed that only 10% actually did (*34*). These observations suggest that mask promotion interventions could be useful in rural areas of low- and middle-income countries, which are home to several billion people at risk for COVID-19.

Results

Our analysis followed our preregistered analysis plan (https://osf.io/vzdh6/) except where indicated. Our primary outcome was symptomatic seroprevalence for SARS-CoV-2. We also analyzed the impact of our intervention on mask-wearing, physical distancing, social distancing, and COVID-19–like symptoms. No adverse events were reported during the study period.

Sample selection

The unions where we conducted our intervention are geographically dispersed throughout rural Bangladesh, as shown in Fig. 1. (Appendix C discusses in more detail how these unions were selected.) Tables S1 and S2 summarize sample selection for our analysis. We initially approved 134,050 households, of which 125,053 provided baseline information. From these 125,053 households, we collected baseline information from 342,183 individuals. Of these, 336,010 (98%) provided symptom data at week 5 and/or 9. Of these, 27,160 (8.0%) reported COVID-19–like symptoms during the 9 weeks since the study began. We attempted to collect blood samples from all symptomatic individuals. Of these, 10,790 (39.7%) consented to have blood collected (40.2% in the treatment group and 39.3% in the control group; $p = 0.24$). We show in table S3 that consent rates are about 40% across men and women and among adults of different age groups in both treatment and control villages.

As such, the sample of individuals for whom we have symptom data is much larger than the sample for whom we have serology data. We tested 9512 (88.2%) of the collected blood samples to determine seroprevalence for SARS-CoV-2 immunoglobulin G (IgG) antibodies. Untested samples (<12%) either lacked sufficient quantity for our test or could not be matched to individuals from our sample because of a barcode scanning error. In our primary outcome analysis, we drop individuals for whom we are missing symptom data or who did not consent to blood sample collection. For the analyses where symptomatic status is the outcome, we report results using both this smaller sample as well as the larger sample of all individuals who provided symptom data. In the baseline, we collected blood samples from a random sample of individuals ($N = 10,085$), of whom 339 had COVID-19–like symptoms. We use these to check balance with respect to baseline symptomatic seropositivity (as well as baseline symptomatic status).

Of the 600 villages initially recruited for the study, the analysis sample excludes four villages where interventions could not be performed owing to a lack of local government cooperation. We exclude an additional 11 villages and their village-pairs (where a village and its village-pair are a control-treatment pair) because we did not observe them in the baseline period before the intervention and one village and its pair for lack of observational data throughout the intervention period, for a total analysis sample of 572 villages.

Primary analyses
Our primary outcomes are balanced at baseline

Although our stratification procedure should have achieved balance with respect to variables observed at the time of randomization, given the many possible opportunities for errors in implementation, we confirm in appendix L that our control and treatment villages are balanced with respect to our primary outcome variables. This assessment was not preregistered. We investigated several other covariates and found a few small imbalances. We checked whether these affect the main results that we report in this paper. For example, we found more 18- to 30-year-olds in the treatment group than in the control group, perhaps because households reported teenagers as 18 years old to receive more masks; our results are robust to dropping this age range.

Our intervention increased mask-wearing

The first column in the top panel of Table 1 reports coefficients from a regression of mask-wearing on a constant, an intervention indicator (based on the assigned groups), baseline mask-wearing, the baseline symptom rate, and indicators for each control-intervention pair. More details of our statistical methods and standard error construction are available in appendix K. Mask-wearing was 13.3% in control villages and 42.3% in treatment villages. Our regression adjusted estimate is an increase of 28.8 percentage points (95% confidence interval = [0.26, 0.31]; numbers in brackets represent 95% confidence intervals throughout the text and tables). If we omit all covariates (except fixed effects for the strata within which we randomized), our point estimate is identical (table S5). Considering only surveillance conducted when no mask distribution was taking place, mask-wearing increased 27.9 percentage points, from 13.4% in control villages to 41.3% in intervention villages (regression adjusted estimate = 0.28 [0.26, 0.30]). We also run our analysis separately in mosques, markets, and other locations such as tea stalls, the entrance of restaurants, and the main road in the village. The increase in mask-wearing was largest in mosques (37.0 percentage points), whereas in all other locations it was 25 to 29 percentage points.

Our intervention increased physical distancing

Contrary to concerns that mask-wearing would promote risk compensation, we did not find evidence that our intervention undermines distancing behavior. In the bottom panel of Table 1, we report identical specifications to the top panel but with physical distancing as the dependent variable. In control villages, 24.1% of observed individuals practiced physical distancing compared with 29.2% in intervention villages, an increase of 5.1% (regression adjusted estimate = 0.05 [0.04, 0.06]). Evidently, protective behaviors like mask-wearing and physical distancing are complements rather than substitutes: Endorsing mask-wearing and informing people about its importance encouraged rural Bangladeshis to take the pandemic more seriously and engage in another form of self-protection. The increases in physical distancing were similar in cloth and surgical mask villages.

Physical distancing increased 5.1 percentage points overall, but there was substantial heterogeneity across locations. In markets, individuals were 7.4 percentage points more likely to physically distance. By contrast, there was no physical distancing practiced in any mosque, in either treatment or control villages, probably as a result of the strong religious norm of standing shoulder-to-shoulder when praying.

Our intervention had no impact on social distancing

It is possible that physical distancing increases because our intervention results in fewer total people being present in public spaces. If socializing increased in the intervention group, but only among risk-conscious people, then we might see physical distancing increase despite people engaging in overall riskier behavior. To assess this, as well as to assess directly if the intervention increased socializing, we studied the effects of our intervention on the total number of people observed at public locations. Although surveillance staff were not able to count everyone in busy public areas, the total number of people they were able to observe gives some indication of the crowd size. We found no difference in the number of people observed in public areas between the treatment and control groups overall (table S6). The social distancing analysis was not preregistered, although the specification exactly parallels our analysis of physical distancing.

Table 1. Mask-wearing and physical distancing, controlling for baseline variables. All regressions include an indicator for each control-intervention pair and baseline symptom rates. The analyses in the top panel control for baseline rates of proper mask-wearing, and the analyses in the bottom panel control for baseline rates of physical distancing. "Baseline symptom rate" is defined as the rate of surveyed individuals in a village who report symptoms coinciding with the WHO definition of a probable COVID-19 case. We assume that (i) all reported symptoms were acute onset, (ii) all people live or work in an area with a high risk of transmission of virus, and (iii) all people have been a contact of a probable or confirmed case of COVID-19 or are linked to a COVID-19 cluster. "No active promotion" refers to any time that surveillance was conducted while promotion was not actively occurring (regardless of the week of the intervention). This excludes surveillance during the Friday Jumma Prayers in the mosque, when promoters were present and actively encouraged mask-wearing. "Other locations" include tea stalls, at the entrance of the restaurant as patrons enter, and the main road to enter the village. "Surgical villages" refer to all treatment villages that received surgical masks as part of the intervention and their control pairs. "Cloth villages" refer to all treatment villages that received cloth masks as part of the intervention and their control pairs. The surgical and cloth subsamples include surveillance from all available locations, equivalent to the column labeled "Full" but run separately for each subgroup. Of the 572 villages included in the analysis sample, we exclude an additional village and its pair in the mosque and market subsamples and two villages and their pairs in the other location subsample because we did not observe them in the baseline period before the intervention. There are 190 treatment villages that received surgical masks as part of the intervention and 96 treatment villages that received cloth masks. Standard errors are in parentheses.

Parameter	Full	No active promotion	Mosques	Markets	Other locations	Surgical mask villages	Cloth mask villages
			Proper mask-wearing				
Intervention coefficient	0.288*** (0.012)	0.279*** (0.011)	0.370*** (0.016)	0.287*** (0.012)	0.251*** (0.012)	0.301*** (0.015)	0.256*** (0.019)
			Physical distancing				
Intervention coefficient	0.051*** (0.005)	0.056*** (0.005)	0.000 (0.000)	0.074*** (0.007)	0.068*** (0.006)	0.054*** (0.006)	0.044*** (0.011)
N villages	572	572	570	570	568	380	192

***Significant at the 1% level.   **Significant at the 5% level.   *Significant at the 10% level.

Our intervention reduced symptomatic seroprevalence

Among the 336,010 participants who completed symptom surveys, 27,160 (8.1%) reported experiencing COVID-19–like illnesses during the study period. More participants in the control villages reported incident COVID-19–like illnesses (N = 13,853; 8.6%) compared with participants in the intervention villages (N = 13,307; 7.6%). More than one-third (39.7%) of symptomatic participants agreed to blood collection. After omitting symptomatic participants who did not consent to blood collection, symptomatic seroprevalence was 0.76% in control villages and 0.68% in the intervention villages. Because the fractions we are reporting omit nonconsenters from the numerator but not the denominator, it is likely that the true rates of symptomatic seroprevalence are substantially higher (perhaps by 2.5 times, if nonconsenters have similar seroprevalence to consenters).

In Table 2 (and table S7), we report results from a regression of symptomatic seroprevalence on a treatment indicator, clustering at the village level and controlling for fixed effects for each pair of control and treatment villages. In the tables, we report results with and without additional controls for baseline symptoms and mask-wearing rates. In table S7, we report results from our prespecified linear model, and in Table 2, we report results from a generalized linear model with a Poisson fam-

ily and log-link function. Here, we discuss the latter results (which are in units of relative risk); the linear model implies results of an almost identical magnitude. The prevalence ratios and accompanying confidence intervals reported in the text correspond to the specifications with baseline controls (hence, "adjusted" prevalence ratio).

The results in all specifications are the same: We estimate a roughly 9% decline in symptomatic seroprevalence in the treatment group (adjusted prevalence ratio = 0.91 [0.82, 1.00]) for a 29 percentage point increase in mask-wearing over 8 weeks. In the second column of Table 2 and table S7, we split our results by mask type (surgical versus cloth). We find clear evidence that surgical masks lead to a relative reduction in symptomatic seroprevalence of 11.1% (adjusted prevalence ratio = 0.89 [0.78, 1.00]; control prevalence = 0.81%; treatment prevalence = 0.72%). Although the point estimates for cloth masks suggests that they reduce risk, the confidence limits include both an effect size similar to surgical masks and no effect at all (adjusted prevalence ratio = 0.94 [0.78, 1.10]; control = 0.67%; treatment = 0.61%).

In appendix N, we investigate the robustness of these results to alternative methods of dealing with missing data from nonconsenters. In the main text, following our prespecified analysis plan, we drop nonconsenting symptomatic individuals. If we instead impute seropositivity for symptomatic nonconsent-

ers based on the population average seropositivity among symptomatic individuals, our pooled estimate of the impact of masking becomes larger and more precise. Notably, with this alternative imputation, we find effects for both cloth and surgical masks on symptomatic seroprevalence.

Not all symptomatic seroprevalence is necessarily a result of infections occurring during our intervention; individuals may have had preexisting SARS-CoV-2 infections and then became symptomatic (perhaps caused by an infection other than SARS-CoV-2). In appendix I, we show that if either (i) masks have the same proportional impact on COVID and non-COVID symptoms or (ii) all symptomatic seropositivity is caused by infections during our intervention, then the percentage decline in symptomatic seroprevalence will exactly equal the decline in symptomatic seroconversions. More generally, the relationship between the two quantities depends on whether masks have a greater impact on COVID or non-COVID symptoms, as well as the proportion of symptomatic seropositivity that is a result of infections preexisting at baseline.

Our intervention reduced WHO COVID-19 symptoms

In Table 3 and table S8, we report results from the same specifications with WHO-defined COVID-19 symptomatic status as the outcome. This is defined as any of following:

Table 2. Symptomatic seroprevalence, expressed in prevalence ratios. All regressions include an indicator for each control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper mask-wearing and baseline symptom rates. "Baseline symptom rate" is defined as the rate of surveyed individuals in a village who report symptoms coinciding with the WHO definition of a probable COVID-19 case. We assume that (i) all reported symptoms were acute onset, (ii) all people live or work in an area with a high risk of transmission of virus, and (iii) all people have been a contact of a probable or confirmed case of COVID-19 or are linked to a COVID-19 cluster. The analysis includes all people surveyed in the baseline household visits, excluding individuals for whom we did not collect midline or endline symptoms, symptomatic individuals from whom we did not collect blood, and individuals from whom we drew blood but did not test their blood. The regressions exclude an additional 17,377 individuals in 34 villages because there are zero people who are symptomatic-seropositive in their village pairs. To check robustness to the type of clustering, in panels 2 and 3 of fig. S2, we show the histogram of effect sizes under "randomization inference" if we randomly reassign treatment within each pair of villages and then estimate our primary specification. We find that our estimated effect size is smaller than 7.0% of the simulated estimates with controls and 7.4% of the simulated estimates without controls (these are the corresponding p values of the randomization inference t test). Blank spaces indicate variables not included in the regression specification reported in each column.

Parameter	Intervention effect	Intervention effect by mask type
No baseline controls		
Intervention prevalence ratio	0.905** [0.815, 0.995]	
Intervention prevalence ratio for surgical mask villages		0.894* [0.782, 1.007]
Intervention prevalence ratio for cloth mask villages		0.925 [0.766, 1.083]
Average symptomatic-seroprevalence rate in paired control villages†	0.0076	0.0076
With baseline controls		
Intervention prevalence ratio	0.905** [0.815, 0.995]	
Intervention prevalence ratio for surgical mask villages		0.889** [0.780, 0.997]
Intervention prevalence ratio for cloth mask villages		0.942 [0.781, 1.103]
N individuals	304,726	304,726
N villages	572	572

***Significant at the 1% level.   **Significant at the 5% level.   *Significant at the 10% level.   †We report the mean rate of symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.

1) Fever and cough.

2) Any three of the following: fever; cough; general weakness and/or fatigue; headache; muscle aches; sore throat; coryza (nasal congestion or runny nose); dyspnoea (shortness of breath or difficulty breathing); anorexia (loss of appetite), nausea, and/or vomiting; diarrhea; or altered mental status.

3) Anosmia (loss of smell) and ageusia (loss of taste).

We find clear evidence that the intervention reduced symptoms: We estimate a reduction of 11.6% (adjusted prevalence ratio = 0.88 [0.83, 0.93]; control = 8.60%; treatment = 7.63%). Additionally, when we look separately by cloth and surgical masks, we find that the intervention led to a reduction in COVID-19–like symptoms under either mask type (p = 0.000 for surgical; p = 0.066 for cloth), but the effect size in surgical mask villages was 30 to 80% larger depending on the specification. In table S9, we run the same specifications using the smaller sample used in our symptomatic seroprevalence regression (i.e., those who consented to give blood). In this sample, we continue to find an effect overall and an effect for surgical masks but see no statistically significant effect for cloth masks.

In-person reinforcement is crucial to our intervention

Our core intervention package combined multiple distinct elements: We provided people with free masks and information about the importance of mask-wearing, we had mask promoters reinforce mask-wearing by stopping individuals in public places who were not wearing masks and reminding them to do so, and we partnered with local leaders to encourage mask-wearing at mosques and markets. Additionally, in some villages, we provided a variety of reminders, commitment devices, and incentives for village leaders. In appendix J, we attempt to disentangle the role played by these different elements in encouraging mask use.

We find no evidence that any of our village-level or household-level treatments, other than mask color, affected mask-wearing. For mask color, we see marginally significant differences that are small in magnitude. In surgical mask villages, blue masks were more likely to be observed than green masks (adjusted percentage point difference = 0.03 [–0.00, 0.06]), and in cloth mask villages, red masks were more likely to be observed than purple masks (adjusted percentage point difference = –0.02

[–0.04, –0.00]). Text message reminders, incentives for village-leaders, or explicit commitment signals explain little of the observed increase in mask-wearing. Compared with self-protection messaging alone, altruistic messaging had no greater impact on mask-wearing, and twice-weekly text messages and a verbal commitment had no significant effects. We saw no significant difference in the rates of mask-wearing in the village-level randomization of surgical versus cloth masks.

We do find nonexperimental evidence that in-person mask promotion and reinforcement is a crucial part of our intervention. Our first pilot study contained all elements of our intervention except in-person reinforcement. Our second pilot study (1 week later) and the full intervention (several months later) added in-person reinforcement. Under the assumption that treatment effects would otherwise be constant over time, we find that in-person reinforcement accounts for 19.2 percentage points of our effect (regression adjusted estimate = 0.19 [–0.33, –0.05]), or 65% of the total effect size. In table S10, we show that this difference is statistically significant whether or not we include baseline controls. This was not a prespecified analysis.
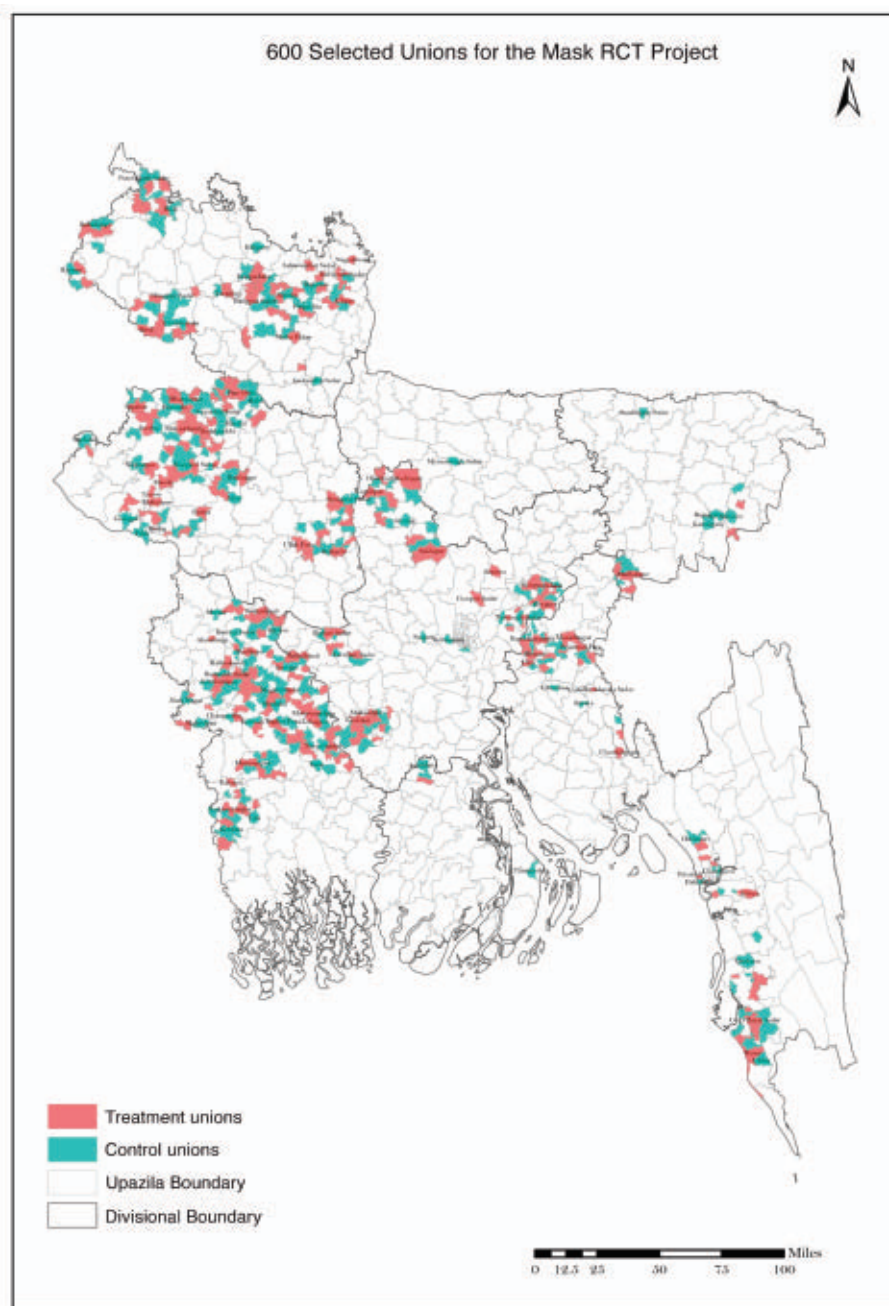
Fig. 1. Map of 600 treatment and control unions. The figure shows the location of the 600 treatment and control unions in the study. RCT, randomized controlled trial; 1 mile = 1.6 km.

Our intervention yields persistent increases in mask-wearing

In appendix M, we present results on mask-wearing after our intervention ended. Even though the door-to-door free mask distribution occurred in the first week only, there was almost no attenuation of mask-wearing over the initial 10 weeks of surveillance. Notably, mask-wearing remained comparably increased in the treatment group during the 2 weeks we continued surveillance after the end of all intervention activities in the village. Three to 4 months later, mask-wearing waned but remained 10 percentage points higher in treatment regions.

Subgroup analyses

Women wear masks more often, but men respond more to the intervention

In table S11, we analyze the impact of our intervention on mask-wearing and physical distancing separately by gender, as well as by whether baseline mask-wearing was above or below the median. Gender was recorded in 65% of observations; age was not recorded during the direct observation surveillance of mask-wearing in public places, and thus we do not conduct an age-stratified assessment. This observed sample is representative of the rural Bangladeshi population that is present in crowded public places during the day; this population is largely composed of men, who have more social contacts outside the home than women. In the gender results, we drop surveillance observations for mosques because in rural Bangladesh it is rare for women to attend mosque. We found that the intervention increased mask-wearing by 27.1 percentage points for men ([0.25, 0.30]) and 22.5 percentage points for women ([0.20, 0.25]). Although we do not have the variation to test this, the gendered difference in effect size may be because our mask promoters were predominantly men or because the mask-wearing rate in control villages was so much higher for women (31% for women versus 12% for men). We intentionally hired predominantly men because most staff interactions would be with men. Men constituted 88.2% of all observed adults. We also found a larger increase in mask-wearing in villages with below-median baseline mask-wearing (where mask-wearing increased from 8.7 to 41.9% at endline) than in those with above-median baseline mask-wearing (where the increase was from 17.5 to 42.6%).

The effect on symptomatic seroprevalence is especially large among the elderly

In Table 4 and table S12, we report results from our primary specification separately by age. Table S12 reports our preregistered specification, a linear model run separately for each decade of age, pooling cloth mask and surgical mask villages. Table 4 synthesizes these results, collapsing by categories of <40, 40 to 49, 50 to 59, and ≥60 years old, reporting results as a relative risk reduction, and showing results separately for surgical and cloth masks. We generally find that the impact of the intervention is concentrated among individuals over age 50. In surgical mask villages, we observe a 22.8% decline in symptomatic seroprevalence among individuals aged 50 to 59 years (adjusted prevalence ratio = 0.77 [0.60, 0.95]) and a 35.3% decline among individuals ≥60 years old in our baseline specification ($p = 0.000$) (adjusted prevalence ratio = 0.65 [0.45, 0.85]). For cloth masks, we find an insignificant (5%) reduction overall but some evidence of a reduction in symptomatic seroprevalence among 40- to 49-year-olds; we investigate more deeply in appendix N and find that the age gradient appears to be sensitive to how we deal with missing values. In the bottom panel of Table 4, we report results where we impute the population average seroprevalence among all nonconsenters rather than dropping them. This alternative approach

Table 3. WHO-defined COVID-19 symptoms, expressed in prevalence ratios. All regressions include an indicator for each control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper mask-wearing and baseline symptom rates. "Baseline symptom rate" is defined as the rate of surveyed individuals in a village who report symptoms coinciding with the WHO definition of a probable COVID-19 case. We assume that (i) all reported symptoms were acute onset, (ii) all people live or work in an area with a high risk of transmission of virus, and (iii) all people have been a contact of a probable or confirmed case of COVID-19 or are linked to a COVID-19 cluster. The analysis includes all people surveyed in the baseline household visits, excluding individuals for whom we did not collect midline or endline symptoms. Blank spaces indicate variables not included in the regression specification reported in each column.

Parameter	Intervention effect	Intervention effect by mask type
No baseline controls		
Intervention prevalence ratio	0.885*** [0.834, 0.934]	
Intervention prevalence ratio for surgical mask villages		0.865*** [0.803, 0.928]
Intervention prevalence ratio for cloth mask villages		0.922* [0.838, 1.005]
Average symptomatic rate in paired control villages†	0.0860	0.0860
With baseline controls		
Intervention prevalence ratio	0.884*** [0.834, 0.934]	
Intervention prevalence ratio for surgical mask villages		0.874*** [0.809, 0.939]
Intervention prevalence ratio for cloth mask villages		0.907** [0.823, 0.991]
N individuals	321,948	321,948
N villages	572	572

***Significant at the 1% level.   **Significant at the 5% level.   *Significant at the 10% level.   †We report the mean rate of symptomatic status at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.

yields more precise overall estimates and suggests that both cloth and surgical masks have greater impacts on symptomatic seroprevalence at older ages, although the impact of surgical masks among those ≥60 years old is smaller than in our baseline specification. Ex ante, it is not obvious to us which imputation method should be preferred, although the second approach makes our results less sensitive to differential consent rates that we observe in some waves of our intervention, as discussed in appendix N.

The effect on WHO COVID-19 symptoms is larger among the elderly

In tables S13 and S14 (the latter being our preregistered specification), we perform the same analysis using the larger sample of individuals who reported symptom information. In this sample, we continue to find larger effects at older ages, although the differences are not as stark as those for the symptomatic seroprevalence outcome. In table S15, we show that the age gradient is steeper for surgical masks.

Men and women have similar reductions in symptoms and symptomatic seroprevalence

In appendix N and table S28, we show results for symptoms and symptomatic seropositivity by gender. We see a similar pattern to the cloth and surgical mask results: We see significant effects for both genders for symptoms and symptomatic seroposivity when we impute seropositivity at the average value for nonconsenters. If we instead drop nonconsenters, the symptomatic seropositivity estimates for men become less precise and are no longer significantly different from zero, whereas the estimates for women remain unchanged.

Additional preregistered specifications

In appendix P, we discuss additional preregistered specifications that are not reported in the text, either because they were substantially underpowered given the available data or because data on required variables were unavailable. We also discuss ways in which trial implementation deviated from our preregistered protocol, such as switching from exclusively phone surveys to household visits at weeks 5 and 9 to increase response rates.

Intervention cost and benefit estimates

In appendix Q, we assess the costs of implementing our intervention relative to the health benefits, specifically focusing on our ongoing efforts to implement this same intervention at scale in Bangladesh. We consider a range of possible estimates for excess deaths from COVID-19 from 1 May to 1 September 2021, and we assume that our age-specific impacts on symptomatic seroprevalence will lead to proportional reductions in mortality. We estimate that a scaled version of our intervention being implemented in Bangladesh will cost about $1.50 per person, and between $10,000 and $52,000 per life saved, depending on which estimate we use for excess deaths.

Discussion

We present results from a cluster-randomized controlled trial of a scalable intervention designed to increase mask-wearing and reduce COVID-19 symptomatic infections. Our estimates suggest that mask-wearing increased by 28.8 percentage points, corresponding to an estimated 51,357 additional adults wearing masks in intervention villages, and this effect was persistent even after active mask promotion was discontinued. The intervention led to a 9.5% reduction in symptomatic SARS-CoV-2 seroprevalence (which corresponds to 105 fewer symptomatic seropositives) and an 11.6% reduction in the prevalence of COVID-19–like symptoms, corresponding to 1541 fewer people reporting these symptoms. If we assume that nonconsenting symptomatic individuals were seropositive at the same rate as consenting symptomatic individuals, the total estimated symptomatic seropositives prevented would be 354. The effects were substantially larger (and more precisely estimated)

Table 4. Symptomatic seroprevalence by age groups and mask type, expressed in prevalence ratios. All regressions include an indicator for each control-intervention pair. The regressions include controls for baseline rates of mask-wearing and baseline symptom rates. "Baseline symptom rate" is defined as the rate of surveyed individuals in a village who report symptoms coinciding with the WHO definition of a probable COVID-19 case. We assume that (i) all reported symptoms were acute onset, (ii) all people live or work in an area with a high risk of transmission of virus, and (iii) all people have been a contact of a probable or confirmed case of COVID-19 or are linked to a COVID-19 cluster. The analysis in the top panel uses the preregistered sample, equivalent to that in Table 2; it includes all people surveyed in the baseline household visits, excluding individuals for whom we did not collect midline or endline symptoms, symptomatic individuals from whom we did not collect blood, and individuals from whom we drew blood but did not test their blood. The analysis in the bottom panel replicates the regressions in the top panel but imputes the seropositivity of individuals from whom we did not draw blood. For symptomatic individuals from whom we did not draw blood, we simulate their symptomatic-seroprevalence status by using the average rate of conditional seropositivity among all symptomatic individuals. This analysis includes all people surveyed in the baseline household visits, excluding individuals for whom we did not collect midline or endline symptoms.

Parameter	All	<40 years old	40 to 49 years old	50 to 59 years old	≥60 years old
Preregistered sample: Drop individuals without blood draws					
Intervention prevalence ratio for surgical mask villages	0.889**	0.967	1.009	0.772**	0.647***
	[0.780, 0.997]	[0.834, 1.100]	[0.817, 1.200]	[0.595, 0.949]	[0.448, 0.845]
Intervention prevalence ratio for cloth mask villages	0.942	1.058	0.713**	0.838	1.084
	[0.781, 1.103]	[0.870, 1.247]	[0.459, 0.967]	[0.524, 1.153]	[0.769, 1.399]
Average symptomatic-seroprevalence in paired control villages†	0.0076	0.0055	0.0095	0.0108	0.0104
N individuals	287,349	146,306	35,839	24,086	27,943
N villages	538	480	384	348	360
Imputing symptomatic-seroprevalence for missing blood draws					
Intervention prevalence ratio for surgical mask villages	0.873***	0.917*	0.975	0.815***	0.701***
	[0.801, 0.945]	[0.829, 1.005]	[0.862, 1.088]	[0.688, 0.942]	[0.577, 0.824]
Intervention prevalence ratio for cloth mask villages	0.890**	0.861***	0.838**	1.153	0.792**
	[0.787, 0.993]	[0.758, 0.965]	[0.678, 0.998]	[0.970, 1.336]	[0.601, 0.983]
Average symptomatic-seroprevalence in paired control villages†	0.0189	0.0152	0.0226	0.0229	0.0251
N individuals	321,383	177,708	51,676	37,340	43,431
N villages	570	566	528	504	534

***Significant at the 1% level.   **Significant at the 5% level.   *Significant at the 10% level.   †We report the mean rate of symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.

in communities where we distributed surgical masks, consistent with their greater filtration efficiency as measured in the laboratory (manuscript forthcoming). In villages randomized to receive surgical masks, the relative reduction in symptomatic seroprevalence was 11% overall, 23% among individuals aged 50 to 59 years, and 35% among those ≥60 years of age in preferred specifications.

We found clear evidence that surgical masks are effective in reducing symptomatic seroprevalence of SARS-CoV-2. Although cloth masks clearly reduce symptoms, we find less clear evidence of their impact on symptomatic SARS-CoV-2 infections, with the statistical significance depending on whether we impute missing values for nonconsenting adults. The number of cloth mask villages (100) was half that for surgical masks (200), meaning that our results tend to be less precise. Additionally, we found evidence that surgical masks were no less likely to be adopted than cloth masks. Surgical masks have higher filtration efficiency, are cheaper, are consistently worn, and are better supported by our evidence as tools to reduce COVID-19 cases.

Our results should not be taken to imply that mask-wearing can prevent only 10% of COVID-19 cases, let alone 10% of COVID-19 mortality. Our intervention induced 29 more people out of every 100 to wear masks, with 42% of people wearing masks in total. The total impact with near-universal masking—perhaps achievable with alternative strategies or stricter enforcement—may be several times larger than our 10% estimate. Additionally, the intervention reduced symptomatic seroprevalence more when surgical masks were used and even more for the highest-risk individuals in our sample (23% for ages 50 to 59 years and 35% for ages ≥60 years). These numbers likely give a better sense of the impact of our intervention on severe morbidity and mortality, because most of the disease burden of the COVID-19 pandemic is borne by the elderly. Where achievable, universal mask adoption is likely to have still larger impacts.

There are several possible theories for why we might observe a larger reduction in COVID-19 cases for older adults. We did not directly measure age during surveillance, but mask-wearing could have increased more for older adults. A second theory is that older adults are more susceptible to infections at viral loads that are preventable by masks. A third theory is that older adults have fewer social connections, so that reducing transmission through any one connection is more likely to prevent infection by severing all transmissible routes. A fourth theory is that people exercised more care and were more likely to wear masks when proximate to the elderly.

We identified a combination of core intervention elements that were effective in increasing mask-wearing in rural Bangladesh: Mask distribution and role-modeling, combined with mask promotion, lead to large and sustained increases in mask use. Results from our pilot studies suggest that combining mask distribution, role-modeling, and active mask promotion—rather than mask distribution and role-modeling alone—seems critical to achieving the full effect. Our trial results also highlight many factors that appear inessential: We find no evidence that public commitments, village-level incentives, text messages, altruistic messaging, or verbal commitments change mask-wearing behavior. The null results on our cross-randomizations do not necessarily imply that these approaches are not worth trying in other contexts, but they teach us that large, persistent increases in mask-wearing are possible without these elements.

Prediction studies that we conducted with policy-makers and public health experts at the WHO and the World Bank before presentations of the study results suggest that our results are informative for policy design. Most of the respondents in the prediction studies anticipated that text messages, verbal commitments, and incentives would increase mask-wearing, when in reality, we estimated fairly precise null effects, and poll respondents believed that in-person mask promotion would have no additional effect, whereas the evidence from our pilot studies suggests that it is essential (for additional details, see appendix R).

Our intervention design is immediately relevant for Bangladesh's plans for larger-scale distribution of masks across all rural areas. The Bangladesh Directorate General of Health has assigned the study team and the nongovernmental organization Bangladesh Rural Advancement Committee (BRAC) the responsibility to scale up the strategies that were proven most effective in this trial to reach 81 million people (35). At the time of writing, we are implementing this program in the 37 districts prioritized by the government based on SARS-CoV-2 test positivity rates. Our results are also relevant for mask dissemination and promotion campaigns planned in other countries and settings that face similar challenges in ensuring mask usage as a result of limited reach and enforcement capacity. The mask promotion model described in this paper was subsequently adopted by governments and other implementers in Pakistan (36), India (37), and Nepal (38). The intervention package would be feasible to implement in a similar fashion in other world regions as well. Beyond face masks, the conceptual underpinning of our strategies could be applied to encourage the adoption of other health behaviors and technologies, in particular, those easily observable by others outside the household, such as purchase and consumption of food, alcohol, and tobacco products in stores, restaurants, or other public spaces (39); hand washing and infection control in health care facilities (40–42); hygiene interventions in childcare and school settings (43, 44); improved sanitation (45, 46); or vaccination drives (47).

Although critics of mask mandates suggest that individuals who wear masks are more likely to engage in high-risk behaviors (48), we found no evidence of risk compensation as a result of increased mask-wearing. Indeed, we found that our intervention slightly increased the likelihood of physical distancing, presumably because individuals participating in the intervention took the threat of COVID-19 more seriously. These findings are consistent with other behaviors, including seat belt use (49) or immunization (50), where risk compensation, even if present, is not sufficient to outweigh direct effects.

The intervention may have influenced rates of COVID-19 by increasing mask use, physical distancing, and/or other risk prevention behaviors. Three factors suggest that the direct impact of masks is the most likely explanation for our documented health impacts. First, in appendix O, we analyze cross-sectionally the relationship between our biological outcomes and both mask-wearing and physical distancing. We find that symptoms and symptomatic seropositivity are negatively correlated with mask-wearing, but not with physical distancing, after controlling for mask-wearing. This analysis uses variation in observational data, rather than solely experimental data, and should therefore be interpreted with caution, as discussed in the appendix. Second, we see no change in physical distancing in the highest-risk environment in our study, typically crowded indoor mosques. However, women do not typically go to mosques in rural Bangladesh, and their symptomatic seropositivity decreased by just as much as that of men, so outdoor transmission or transmission in settings that we do not observe directly may be important. Third, our study complements a large body of laboratory and quasi-experimental evidence that masks have a direct effect on SARS-CoV-2 transmission (1).

We estimate that a scaled version of our intervention being implemented in Bangladesh will cost between $10,000 and $52,000 per life saved, depending on what fraction of excess deaths are attributable to COVID-19. This is considerably lower than the value of a statistical life in Bangladesh [$205,000 (51)] and, under severe outbreaks, is comparable to the most cost-efficient humanitarian programs at scale [e.g., distributing insecticide nets to prevent malaria costs $9200 per life saved (52)]. This estimate includes only mortality impacts and not morbidity, and greater cost-efficiency is possible if our intervention can be streamlined to further isolate the essential components. Most of our costs were the personnel costs for mask-promoters: If we consider only the costs of mask production, these numbers would be 20 times lower. Thus, the overall cost to save a life in countries where mask mandates can be enforced at minimal cost with existing infrastructure may be substantially lower than our estimates above.

Study limitations

Our study has several limitations. The distinct appearance of project-associated masks and increased mask-wearing in intervention villages made it impossible to blind surveillance staff to study-arm assignment. However, staff were not informed about the exact purpose of the study. Even though surveillance staff were plain-clothed and were instructed to remain discreet, community members could have recognized that they were being observed and changed their behavior. Additionally, survey respondents could have changed their likelihood of reporting symptoms in places where mask-wearing was more widespread. If respondents were more cognizant of symptoms in mask-wearing areas, this may bias us toward underestimating the impact of masks; if respondents in mask-wearing areas were less concerned with mild symptoms and thus were less likely to recall them, this might bias us toward overestimating the impact of masks. Although we confirm that blood consent rates are not significantly different in the treatment and control groups and are comparable across all demographic groups, we cannot rule out that the composition of consenters differed between the treatment and control groups. The slightly higher point estimate for consent in the treatment group biases us away from finding an effect, because it raises symptomatic seroprevalence in the treatment group. Although control villages were at least 2 km from intervention villages, adults from control villages may have come to intervention villages to receive masks, reducing the apparent impact of the intervention. Although we did not directly assess harms in this study, there could be costs resulting from discomfort with increased mask-wearing, adverse health effects such as dermatitis or headaches, or impaired communication.

Because the study was powered to detect differences in symptomatic seroprevalence, we cannot distinguish whether masks work by making symptoms less severe (through a reduced viral load at transmission) or by reducing new infections. We selected the WHO case definition of COVID-19 for its sensitivity, though its limited specificity may imply that the impact of masks on symptoms comes partly from non–SARS-CoV-2 respiratory infections. If masks reduce COVID-19 by reducing symptoms (for a given number of infections), they could help ease the morbidity and mortality resulting from a given number of SARS-CoV-2 infections. If masks reduce infections, they may reduce the total number of infections over the long-term by buying more time to increase the fraction of the population that is vaccinated. At the time of the study, the predominant circulating SARS-CoV-2 strain was B.1.1.7 (Alpha) (53). The impacts of the Delta variant on the number of infections prevented by a given mask-wearer are uncertain; the population-wide consequences of infections prevented by a given mask-wearer may be larger given a higher reproduction number.

We found that mask distribution, role modeling, and promotion in a low- and middle-income country setting increased mask-wearing and physical distancing, leading to lower illness, particularly in older adults. We find especially robust evidence that surgical masks prevent COVID-19. Whether people with respiratory symptoms should generally wear masks to

prevent respiratory virus transmission, including for viruses other than SARS-CoV-2, is an important area for future research. Our findings suggest that such behavior may benefit public health.

Methods and materials
Sampling frame and timeline

The intervention protocol, prespecified analysis plan, and CONSORT checklist are available at https://osf.io/vzdh6/. We discuss our sample-size calculations in appendix B and discuss the selection and pairwise randomization in appendix C. In brief, we stratified villages based on geographic location and available case data, and then selected one treatment and one control village from each pair.

Village-level cluster randomization was important for three reasons. First, unlike technologies with primarily private benefits, mask adoption is likely to yield especially large benefits at the community level. Second, mask adoption by some may influence mask adoption by others because mask-wearing is immediately visible to other members of the community (45). Third, this design allows us to assess the full impact of masks on symptomatic infections, including through source control. Individual-level randomization would identify only whether masks protect wearers.

Our intervention was designed to last 8 weeks in each village. The intervention started in different villages at different times, rolling out over a 6-week period in seven waves. There were between 16 and 61 village-pairs grouped in each wave based on geographic proximity, and paired control and treatment villages were always included in the same wave. The first wave was rolled out on 17 and 18 November 2020 and the last wave was rolled out on 5 and 6 January 2021.

Innovations for Poverty Action (IPA) staff traveled to many villages that had low mask uptake in the first 5 weeks of the study and found that in these villages, local leaders were not very engaged in supporting mask promotion. Hence, we retrained mask-promotion staff partway through the intervention to work more closely with local leaders and set specific milestones for that partnership.

Outcomes

Our primary outcome was symptomatic seroprevalence of SARS-CoV-2. Our secondary outcomes were prevalence of proper mask-wearing, physical distancing, and symptoms consistent with COVID-19. For COVID-19 symptoms, we used the symptoms that correspond to the WHO case definition of probable COVID-19 given epidemiological risk factors: (i) fever and cough; (ii) three or more of the following symptoms (fever; cough; general weakness and/or fatigue; headache; myalgia; sore throat;

coryza; dyspnea; anorexia, nausea, and/or vomiting; diarrhea; and altered mental status); or (iii) loss of taste or smell. Seropositivity was defined by having detectable IgG antibodies against SARS-CoV-2.

Intervention materials and activities

Our entire intervention was designed to be easily adopted by other nongovernmental organizations or government agencies and required minimal monitoring. We have made the materials public in multiple languages to ease widespread adoption and replication by other implementers (https://osf.io/23mws/).

We provide design specifications for our masks in appendix F. We used high-quality surgical masks that had a filtration efficiency of 95% [standard deviation (SD) = 1%]; this is substantially higher than the filtration efficiency of the cloth masks we designed, which had a filtration efficiency of 37% (SD = 6%). These cloth masks had substantially higher filtration than common commercial three-ply cotton masks but lower filtration than hybrid masks that use materials not commonly available for community members in low-resource settings (54). Although cloth masks have less leakage because they fit the face more closely (55) and can be sewn without specialized equipment, they are an order of magnitude more expensive than surgical masks. The filtration efficiency of the high-quality surgical masks used in this study was 76% after washing them with bar soap and water 10 times (manuscript forthcoming). Although surgical masks can break down into microplastics that can enter the environment if disposed of improperly, an analysis of waste generated in Bangladesh's first lockdown finds that the mass of surgical mask waste was one-third that of polyethylene bags, which also break down into macro- and microplastics (56–58).

Surgical masks were outfitted with a sticker that had a logo of a mask with an outline of the Bangladeshi flag and a phrase in Bengali that noted that the mask could be washed and reused (59). The relatively large scale of our bulk order allowed us to negotiate mask prices of $0.50 per cloth mask and $0.13 per surgical mask ($0.06 of which was the cost of a sticker reminding people that they could wash and reuse the surgical mask).

Adult household members were asked to wear masks whenever they were outside their house and around other people. To emphasize the importance of mask-wearing, we prepared a brief video of notable public figures discussing why, how, and when to wear a mask. The video was shown to each household during the mask distribution visit and featured the Honorable Prime Minister of Bangladesh Sheikh Hasina, the head of the Imam Training Academy, and the national cricket star Shakib Al Hasan. During the distribution

visit, households also received a brochure based on WHO materials that depicted proper mask-wearing.

We implemented a basic set of interventions in all treatment villages and cross-randomized additional intervention elements in randomly chosen subsets of treatment villages to investigate whether those have any additional impact on mask-wearing. The basic intervention package consists of five main elements:

1) One-time mask distribution and information provision (about masks) at households.

2) Mask distribution in markets for 3 to 6 days per week during all 8 weeks of the intervention.

3) Mask distribution at mosques on three Fridays during the first 4 weeks of the intervention.

4) Mask promotion in public spaces and markets where non–mask wearers were encouraged to wear masks (weekly or biweekly).

5) Role modeling and advocacy by local leaders, including imams discussing the importance of mask-wearing at Friday prayers using a scripted speech provided by the research team.

Participants and mask surveillance staff were not told which villages were in which intervention arm, but the intervention materials were clearly visible. The prespecified analyses and sample exclusions were made by analysts blinded to the treatment assignment.

Cross-randomization of behavior change communication and incentives
Village-level cross-randomizations

Within the intervention arm, we cross-randomized villages to four village-level and four household-level treatments to test the impact of a range of social and behavior change communication strategies on mask-wearing. All intervention villages were assigned to either the treatment or the control group of each of these four randomizations. These village-level randomizations were as follows:

1) Randomization of treated villages to either cloth or surgical masks.

2. Randomization of treated villages to public commitment (providing households signage and asking them to place signage on doors that declares they are a mask-wearing household) or not. The signage was meant to encourage formation of social norms through public signaling.

3. Randomization of treated villages to no incentive, nonmonetary incentive, or monetary incentive of $190 given to the village leader for a project benefitting the public. We announced that the monetary reward or the certificate would be awarded if village-level mask-wearing among adults exceeded 75% at 8 weeks after the intervention started.

4. Randomization of treated villages to 0 or 100% of households receiving twice-weekly

text message reminders about the importance of mask-wearing.

Household-level cross-randomizations

We had three household-level cross-randomizations. In any single village, only one of these household randomizations was operative. Because our data collection protocols relied on passive observation at the village level, we could not record the mask-wearing behavior of individual households. To infer the effect of the household-level treatments, we therefore varied the color of the masks distributed to the household based on its cross-randomization status and had surveillance staff record the mask color of observed individuals. In surgical mask villages, a household received blue or green masks and promoters distributed an equal number of blue and green masks in public settings. In cloth mask villages, households received violet or red masks and promoters distributed blue masks in public settings. To avoid conflating the effect of the household-specific treatment with the effect of the mask color, we randomized which color corresponded to which treatment status across villages (this way a specific color was not fully coincident with a specific treatment). The household-level randomizations, described in further detail in appendix D and visualized in fig. S1, were as follows:

1) Households were randomized to receive messages emphasizing either altruism or self-protection.

2) Households were randomized to making a verbal commitment to be a mask-wearing household (all adults in the household promise to wear a mask when they are outside and around other people) or not. This experiment was conducted in a third set of villages where there was no public signage commitment.

3) Households were randomized to receive twice-weekly text reminders or not. As mentioned above, the text message saturation was randomly varied to 0, 50, or 100% of all households receiving texts, and in the 50% villages, the specific households that received the texts was also random.

Conceptual basis for tested social and behavioral change communication

We selected intervention elements that had a reasonable chance of persuading rural Bangladeshis to wear masks by consulting literature in public health, development and behavioral economics, and marketing to identify some of the most promising strategies. An extensive literature identifies price and access as key deterrents to the adoption of welfare-improving products, and especially of technologies that produce positive health externalities, such as face masks (21, 60). Household distribution of free face masks therefore formed the core part of our strat-

egy. Inspired by large literature in marketing and economics on the role of opinion leaders in new product diffusion, we additionally emphasized a partnership with community leaders in mask distribution (25, 61).

The additional village- and household-level treatments we experimented with were also motivated by insights from marketing, public health, development, and behavioral economics. For example, masks are a visible good where social norms are expected to be important, so we consulted the literature that documented peer effects in product adoption (62–65). We experimented with incentives because it is unclear whether extrinsic rewards crowd out intrinsic motivation (66–68). We tested whether soft commitment devices encourage targets to follow through with actual behavior change (69, 70), whether public displays can promote social norms (27), whether an altruistic framing inspires people more or less than self-interest (71), whether social image concerns and signaling can lead to higher compliance (22, 72), and whether regular reminders are a useful tool to ensure adoption (23).

Piloting interventions

IPA implemented two pilot studies: Pilot 1 from 22 to 31 July 2020 and Pilot 2 from 13 to 26 August 2020. The objective of the pilot studies was to mimic some of the major aspects of the main experiment to identify implementation challenges. Each pilot study was conducted in 10 unions that were not part of the main study area. We used the difference between the pilot studies to better understand which elements of our full intervention were essential. We also conducted focus group discussions and in-depth interviews with village residents, community leaders, religious leaders, and political leaders to elicit opinions on how to maximize the effectiveness of the intervention.

Surveillance strategies

Mask-wearing and physical distancing were measured through direct observation. Surveillance was conducted using a standard protocol that instructed staff to spend 1 hour at each of the following high-traffic locations in the village: market, restaurant entrances, main road, tea stalls, and mosque; the location and timing changed so that the mask-wearing and physical distancing practices of as many individuals as possible could be recorded. Although SARS-CoV-2 transmission is more likely in indoor locations with limited ventilation than outside, rural Bangladeshi villages have few nonresidential spaces where people gather, so observations were conducted outside except at the mosque, where surveillance was conducted inside.

Surveillance staff were distinct from intervention implementation staff and conducted

surveillance in paired intervention and control villages. To minimize the likelihood that village residents would perceive that their mask-wearing behavior was being observed, surveillance staff were separate from mask promoters and wore no identifying apparel while passively observing mask-wearing and physical distancing practices in the communities. They recorded the mask-wearing behavior of all of the adults that they were able to observe during surveillance periods; observations were not limited to adults from enrolled households. Surveillance staff noted whether adults were wearing any mask or face covering, whether the mask was one distributed by our project (and, if so, the color), and how the mask was worn. We defined proper mask-wearing as wearing either a project mask or an alternative face-covering over the mouth and nose and improper mask-wearing as wearing a mask in any way that did not fully cover the mouth and nose. Surveillance staff observed a single individual and recorded that person as practicing physical distancing if he or she was at least one arm's length away from all other people. Additional details are available in appendix G.

Symptomatic SARS-CoV-2 testing
Symptom reporting

The owner of the household's primary phone completed surveys by phone or in-person at weeks 5 and 9 after the start of the intervention. They were asked to report symptoms experienced by any household member that occurred in the previous week and over the previous month. COVID-19–like symptoms were defined by whether they were consistent with the WHO COVID-19 case definition for suspected or probable cases with an epidemiological link (73).

Blood sample collection

We collected endline capillary blood samples from participants who reported COVID-19–like symptoms during the study period and consented to blood collection. We additionally collected samples on a subset of randomly selected participants at baseline, independent of symptoms, to assess overall seropositivity. For the purposes of blood collection, endline was defined as 10 to 12 weeks from the start of the intervention. Blood samples were obtained by puncture with a 20-gauge safety lancet to the third or fourth digit. Five hundred microliters of blood were collected into Microtainer capillary blood collection serum separator tubes (BD, Franklin Lakes, NJ). Blood samples were transported on ice and stored at −20°C until testing.

SARS-CoV-2 testing

Blood samples were tested for the presence of IgG antibodies against SARS-CoV-2 using the

SCoV-2 Detect IgG ELISA kit (InBios, Seattle, WA). This assay detects IgG antibodies against the spike protein subunit (S1) of SARS-CoV-2. The assays were performed according to the manufacturer's instructions. Additional details are presented in appendix H.

Symptomatic seropositivity

Our primary outcome is symptomatic seropositivity. As noted above, individuals are symptomatic if they (i) meet the WHO surveillance definition of probable COVID-19 illness and (ii) are seropositive in our blood test at end-line. If either of these conditions fail to hold, $Y_{ij} = 0$, where Y_{ij} is an indicator for whether individual i in village j is symptomatic seropositive. To assess seropositivity, we tested all individuals who were symptomatic in either our 5- or 9-week household survey.

Our goal is to estimate the impact of the intervention on symptomatic seropositivity, defined as $\psi_0 = E_x[E(Y_{ij}|T_j = 1, x_j) - E(Y_{ij}|T_j = 0, x_j)]$ where T_j is an indicator for whether a village was treated and x_j are village-level covariates, including baseline mask use in each village (constructed as described below) and baseline influenza-like illness and COVID-19 illness based on reported symptoms, as well as indicators for each pair of villages from our pairwise stratification method.

In our preregistered specification, we estimate this parameter by ordinary least squares, clustering at the village level using the approach in (74–76). The dependent variable is Y_{ij}, the independent variable of interest is T_j, and controls are included for the x_j covariates, including baseline mask use and baseline respiratory symptom rates in each village. We also report results from a generalized linear model with a Poisson family and log-link function to compute relative risk (77). More details of our statistical analyses are reported in appendix K.

REFERENCES AND NOTES

1. J. Howard et al., An evidence review of face masks against COVID-19. Proc. Natl. Acad. Sci. U.S.A. 118, e2014564118 (2021). doi: 10.1073/pnas.2014564118; pmid: 33431650
2. N. H. L. Leung et al., Respiratory virus shedding in exhaled breath and efficacy of face masks. Nat. Med. 26, 676–680 (2020). doi: 10.1038/s41591-020-0843-2; pmid: 32371934
3. C. R. MacIntyre, A. A. Chughtai, Facemasks for the prevention of infection in healthcare and community settings. BMJ 350, h694 (2015). doi: 10.1136/bmj.h694; pmid: 25858901
4. H. Bundgaard et al., Effectiveness of adding a mask recommendation to other public health measures to prevent SARS-CoV-2 infection in Danish mask wearers: A randomized controlled trial. Ann. Intern. Med. 174, 335–343 (2021). doi: 10.7326/M20-6817; pmid: 33205991
5. C. N. Ngonghala et al., Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel Coronavirus. Math. Biosci. 325, 108364 (2020). doi: 10.1016/j.mbs.2020.108364; pmid: 32360770
6. C. T. Leffler et al., Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks. Am. J. Trop. Med. Hyg. 103, 2400–2411 (2020). doi: 10.4269/ajtmh.20-1015; pmid: 33124541
7. W. Lyu, G. L. Wehby, Community use of face masks and COVID-19: Evidence from a natural experiment of state mandates in the US. Health Aff. 39, 1419–1425 (2020). doi: 10.1377/hlthaff.2020.00818; pmid: 32543923
8. V. Chernozhukov, H. Kasahara, P. Schrimpf, Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. J. Econom. 220, 23–62 (2021). doi: 10.1016/j.jeconom.2020.09.003; pmid: 33100476
9. J. Abaluck et al., The case for universal cloth mask adoption and policies to increase supply of medical masks for health workers. SSRN 3567438 [Preprint] (2020); doi: 10.2139/ssrn.3567438
10. Y. Cheng et al., Face masks effectively limit the probability of SARS-CoV-2 transmission. Science 372, 1439–1443 (2021). doi: 10.1126/science.abg0242; pmid: 34016743
11. A. Mullard, "How COVID vaccines are being divvied up around the world," Nature News, 30 November 2020.
12. T. A. Ghebreyesus, WHO Director-General's opening remarks at the media briefing on COVID-19 - 5 June 2020 (World Health Organization, 2020).
13. L. M. Brosseau, M. Sietsema, "Commentary: Masks-for-all for COVID-19 not based on sound data" (Center for Infectious Disease Research and Policy, University of Minnesota, 2020); www.cidrap.umn.edu/news-perspective/2020/04/commentary-masks-all-covid-19-not-based-sound-data.
14. M. A. Johansson et al., SARS-CoV-2 transmission from people without COVID-19 symptoms. JAMA Netw. Open 4, e2035057 (2021). doi: 10.1001/jamanetworkopen.2020.35057; pmid: 33410879
15. Centers for Disease Control and Prevention (CDC), "Science brief: Community use of cloth masks to control the spread of SARS-CoV-2" (CDC, 2021).
16. J. M. Brophy, "Covid-19: Controversial trial may actually show that masks protect the wearer," BMJ Opinion, 24 November 2020.
17. J. Pan, C. Harb, W. Leng, L. C. Marr, Inward and outward effectiveness of cloth masks, a surgical mask, and a face shield. Aerosol Sci. Technol. 55, 718–733 (2021). doi: 10.1080/02786826.2021.1890687
18. D. Kahneman, D. T. Miller, Norm theory: Comparing reality to its alternatives. Psychol. Rev. 93, 136–153 (1986). doi: 10.1037/0033-295X.93.2.136
19. J. Jordan, E. Yoeli, D. Rand, Don't get it or don't spread it: Comparing self-interested versus prosocial motivations for COVID-19 prevention behaviors. PsyArXiv [Preprint] (2020); doi: 10.31234/osf.io/yuq7x
20. R. B. Cialdini, N. J. Goldstein, Social influence: Compliance and conformity. Annu. Rev. Psychol. 55, 591–621 (2004). doi: 10.1146/annurev.psych.55.090902.142015; pmid: 14744228
21. M. Bates, R. Glennerster, K. Gumede, E. Duflo, The price is wrong. Field Actions Sci. Rep. 4, 30 (2012).
22. A. Karing, "Social signaling and childhood immunization: A field experiment in Sierra Leone," Working paper, University of California, Berkeley (2018).
23. D. Karlan, M. McConnell, S. Mullainathan, J. Zinman, Getting to the top of mind: How reminders increase saving. Manage. Sci. 62, 3393–3411 (2016). doi: 10.1287/mnsc.2015.2296
24. N. J. Goldstein, R. B. Cialdini, V. Griskevicius, A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. J. Consum. Res. 35, 472–482 (2008). doi: 10.1086/586910
25. G. Miller, A. M. Mobarak, Learning about new technologies through social networks: Experimental evidence on nontraditional stoves in Bangladesh. Mark. Sci. 34, 480–499 (2014). doi: 10.1287/mksc.2014.0845
26. P. Manchanda, Y. Xie, N. Youn, The role of targeted communication and contagion in product adoption. Mark. Sci. 27, 961–976 (2008). doi: 10.1287/mksc.1070.0354
27. C. Bicchieri, Norms in the Wild: How to Diagnose, Measure, and Change Social Norms (Oxford Univ. Press, 2016).
28. T. R. Bhuiyan et al., SARS-CoV-2 seroprevalence in Chattogram, Bangladesh before the Delta surge, March-June 2021. medRxiv 2021.07.16.21260611 [Preprint] (2021); doi: 10.1101/2021.07.16.21260611
29. icddr,b, "Higher covid-19 seropositivity observed among residents in Dhaka and Chattogram," 22 June 2021; www.icddrb.org/news-and-events/news?id=878.
30. Management Information System (MIS), Directorate General of Health Services (DGHS), COVID-19 dynamic dashboard for Bangladesh (2021); http://dashboard.dghs.gov.bd/webportal/pages/covid19.php [accessed 16 August 2021].
31. M. V. Murhekar et al., SARS-CoV-2 seroprevalence among the general population and healthcare workers in India, December 2020-January 2021. Int. J. Infect. Dis. 108, 145–155 (2021). doi: 10.1016/j.ijid.2021.05.040; pmid: 34022338
32. A. Anand, J. Sandefur, A. Subramanian, "Three new estimates of India's all-cause excess mortality during the COVID-19 pandemic," Working paper no. 589, Center for Global Development, July 2021.
33. J. Abaluck, A. M. Mobarak, "Getting all Bangladeshis to wear masks," WhiteBoard Magazine, 1 December 2020.
34. A. Jakubowski et al., Self-reported mask wearing greatly exceeds directly observed use: Urgent need for policy intervention in Kenya. medRxiv 2021.01.27.21250487 [Preprint] (2021); .doi: 10.1101/2021.01.27.21250487
35. K. K. Tithila, "Brac's efforts to mask up Bangladesh could be game-changer," Dhaka Tribune (Bangladesh), 15 July 2021.
36. S. Riaz, "Punjab authorities kick off 'NORM' campaign to increase mask-wearing," Arab New (Pakistan), 30 June 2021.
37. S. Bhattacharjee, "Covid-19 Crisis: India draws lessons from Bangladesh's mask study," Business Standard (India), 15 May 2021.
38. Republica, "Nepal Mask Campaign launches with the slogan 'Let's wear masks, let's save each other's lives,'" myRepública, 16 August 2021.
39. G. J. Hollands et al., Altering the availability or proximity of food, alcohol, and tobacco products to change their selection and consumption. Cochrane Database Syst. Rev. 2019, CD012576 (2019). doi: 10.1002/14651858.CD012573.pub3; pmid: 31482606
40. S. Naikoba, A. Hayward, The effectiveness of interventions aimed at increasing handwashing in healthcare workers – a systematic review. J. Hosp. Infect. 47, 173–180 (2001). doi: 10.1053/jhin.2000.0882; pmid: 11247676
41. C. Houghton et al., Barriers and facilitators to healthcare workers' adherence with infection prevention and control (IPC) guidelines for respiratory infectious diseases: A rapid qualitative evidence synthesis. Cochrane Database Syst. Rev. 4, CD013582 (2020). doi: 10.1002/14651858.CD013582; pmid: 32315451
42. H. J. Seo et al., Interventions to improve hand hygiene compliance in emergency departments: A systematic review. J. Hosp. Infect. 102, 394–406 (2019). doi: 10.1016/j.jhin.2019.03.013; pmid: 30935982
43. D. Biswas et al., Effectiveness of a behavior change intervention with hand sanitizer use and respiratory hygiene in reducing laboratory-confirmed influenza among schoolchildren in Bangladesh: A cluster randomized controlled trial. Am. J. Trop. Med. Hyg. 101, 1446–1455 (2019). doi: 10.4269/ajtmh.19-0376; pmid: 31701861
44. S. L. McGuinness et al., Effect of hygiene interventions on acute respiratory infections in childcare, school and domestic settings in low- and middle-income countries: A systematic review. Trop. Med. Int. Health 23, 816–833 (2018). doi: 10.1111/tmi.13080; pmid: 29799658
45. R. Guiteras, J. Levinsohn, A. M. Mobarak, Sanitation subsidies. Encouraging sanitation investment in the developing world: A cluster-randomized trial. Science 348, 903–906 (2015). doi: 10.1126/science.aaa0491; pmid: 25883316
46. S. R. Patil et al., The effect of India's total sanitation campaign on defecation behaviors and child health in rural Madhya Pradesh: A cluster randomized controlled trial. PLOS Med. 11, e1001709 (2014). doi: 10.1371/journal.pmed.1001709; pmid: 25157929
47. J. S. Solís Arce et al., COVID-19 vaccine acceptance and hesitancy in low- and middle-income countries. Nat. Med. 27, 1385–1394 (2021). doi: 10.1038/s41591-021-01454-y; pmid: 34272499
48. Y. Yan, J. Bayham, A. Richter, E. P. Fenichel, Risk compensation and face mask mandates during the COVID-19 pandemic. Sci. Rep. 11, 3174 (2021). doi: 10.1038/s41598-020-79139-8; pmid: 33414495
49. A. Cohen, L. Einav, Estimating risk preferences from deductible choice. Am. Econ. Rev. 97, 745–788 (2007). doi: 10.1257/aer.97.3.745
50. M. L. Kasting, G. K. Shapiro, Z. Rosberger, J. A. Kahn, G. D. Zimet, Tempest in a teapot: A systematic review of HPV vaccination and risk compensation research. Hum. Vaccin. Immunother. 12, 1435–1450 (2016). doi: 10.1080/21645515.2016.1141158; pmid: 26864126
51. W. K. Viscusi, C. J. Masterman, Income elasticities and global values of a statistical life. J. Benefit Cost Anal. 8, 226–250 (2017). doi: 10.1017/bca.2017.12
52. GiveWell, 2021 GiveWell cost-effectiveness analysis – Version 1 (2021); www.givewell.org/how-we-work/our-criteria/cost-effectiveness/cost-effectiveness-models/changelog-2021#Version_1_Published_May_5_2021 [accessed 4 January 2021].
53. J. Hadfield et al., Nextstrain: Real-time tracking of pathogen evolution. Bioinformatics 34, 4121–4123 (2018). doi: 10.1093/bioinformatics/bty407; pmid: 29790939

54. L. H. Kwong et al., Review of the breathability and filtration efficiency of common household materials for face masks. ACS Nano 15, 5904–5924 (2021). doi: 10.1021/acsnano.0c10146; pmid: 33822580

55. S. Duncan, P. Bodurtha, S. Naqvi, The protective performance of reusable cloth face masks, disposable procedure masks, KN95 masks and N95 respirators: Filtration and total inward leakage. PLOS ONE 16, e0258191 (2021). doi: 10.1371/journal.pone.0258191; pmid: 34614026

56. O. O. Fadare, E. D. Okoffo, Covid-19 face masks: A potential source of microplastic fibers in the environment. Sci. Total Environ. 737, 140279 (2020). doi: 10.1016/j.scitotenv.2020.140279; pmid: 32563114

57. Environment and Social Development Organization (ESDO), "COVID-19 pandemic pushes single use plastic waste outbreak: No management, no protection: High health and environmental risk unveil" (ESDO, 2020).

58. I. M. Steensgaard et al., From macro- to microplastics – Analysis of EU regulation along the life cycle of plastic bags. Environ. Pollut. 224, 289–299 (2017). doi: 10.1016/j.envpol.2017.02.007; pmid: 28222979

59. E.-S. Jang, C.-W. Kang, Do face masks become worthless after only one use in the COVID-19 pandemic? Infect. Chemother. 52, 583–591 (2020). doi: 10.3947/ic.2020.52.4.583; pmid: 33263241

60. M. Kremer, E. Miguel, The illusion of sustainability. Q. J. Econ. 122, 1007–1065 (2007). doi: 10.1162/qjec.122.3.1007

61. P. S. van Eck, W. Jager, P. S. Leeflang, Opinion leaders' role in innovation diffusion: A simulation study. J. Prod. Innov. Manage. 28, 187–203 (2011). doi: 10.1111/j.1540-5885.2011.00791.x

62. E. Oster, R. Thornton, Determinants of technology adoption: Peer effects in menstrual cup take-up. J. Eur. Econ. Assoc. 10, 1263–1293 (2012). doi: 10.1111/j.1542-4774.2012.01090.x

63. H. Allcott, Social norms and energy conservation. J. Public Econ. 95, 1082–1095 (2011). doi: 10.1016/j.jpubeco.2011.03.003

64. R. Guiteras, J. Levinsohn, A. M. Mobarak, "Demand estimation with strategic complementarities: Sanitation in Bangladesh," Discussion paper no. DP13498, Centre for Economic Policy Research, January 2019.

65. L. Beaman, A. BenYishay, J. Magruder, A. M. Mobarak, Can network theory-based targeting increase technology adoption? Am. Econ. Rev. 111, 1918–1943 (2021). doi: 10.1257/aer.20200295

66. N. Ashraf, O. Bandiera, K. Jack, No margin, no mission? A field experiment on incentives for public service delivery. J. Public Econ. 120, 1–17 (2014). doi: 10.1016/j.jpubeco.2014.06.014

67. R. Chetty, E. Saez, L. Sandor, What policies increase prosocial behavior? An experiment with referees at the Journal of Public Economics. J. Econ. Perspect. 28, 169–188 (2014). doi: 10.1257/jep.28.3.169

68. D. Ariely, A. Bracha, S. Meier, Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. Am. Econ. Rev. 99, 544–555 (2009). doi: 10.1257/aer.99.1.544

69. G. Bryan, D. Karlan, S. Nelson, Commitment devices. Annu. Rev. Econ. 2, 671–698 (2010). doi: 10.1146/annurev.economics.102308.124324

70. J. Luoto, D. Levine, J. Albert, S. Luby, Nudging to use: Achieving safe water behaviors in Kenya and Bangladesh. J. Dev. Econ. 110, 13–21 (2014). doi: 10.1016/j.jdeveco.2014.02.010

71. N. Ashraf, O. Bandiera, E. Davenport, S. S. Lee, Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. Am. Econ. Rev. 110, 1355–1394 (2020). doi: 10.1257/aer.20180326

72. L. Bursztyn, R. Jensen, Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. Annu. Rev. Econ. 9, 131–153 (2017). doi: 10.1146/annurev-economics-063016-103625

73. World Health Organization (WHO), WHO COVID-19 case definition (2020); www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance_Case_Definition-2020.2 [accessed on 15 October 2020].

74. P. Guimarães, P. Portugal, A simple feasible procedure to fit models with high-dimensional fixed effects. Stata J. 10, 628–649 (2010). doi: 10.1177/1536867X1101000406

75. S. Gaure, "OLS with multiple high dimensional category dummies," Memorandum 14/2010, Oslo University (2011).

76. P. Guimarães, P. Portugal, A simple feasible procedure to fit models with high-dimensional fixed effects. Stata J. 10, 628–649 (2011). doi: 10.1177/1536867X1101000406

77. G. Zou, A modified poisson regression approach to prospective studies with binary data. Am. J. Epidemiol. 159, 702–706 (2004). doi: 10.1093/aje/kwh090; pmid: 15033648

78. J. Abaluck et al., Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh. Zenodo (2021); doi: 10.5281/zenodo.5703876

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi9069
Figs. S1 to S6
Tables S1 to S37
Appendices
References (79–81)
MDAR Reproducibility Checklist

RESEARCH ARTICLES

CORONAVIRUS

4′-Fluorouridine is an oral antiviral that blocks respiratory syncytial virus and SARS-CoV-2 replication

Julien Sourimant[1], Carolin M. Lieber[1], Megha Aggarwal[1], Robert M. Cox[1], Josef D. Wolf[1], Jeong-Joong Yoon[1], Mart Toots[1], Chengin Ye[2], Zachary Sticher[3], Alexander A. Kolykhalov[3,4], Luis Martinez-Sobrido[2], Gregory R. Bluemling[3,4], Michael G. Natchus[3], George R. Painter[3,4,5], Richard K. Plemper[1,6]*

The COVID-19 pandemic has underscored the critical need for broad-spectrum therapeutics against respiratory viruses. Respiratory syncytial virus (RSV) is a major threat to pediatric patients and older adults. We describe 4′-fluorouridine (4′-FIU, EIDD-2749), a ribonucleoside analog that inhibits RSV, related RNA viruses, and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with high selectivity index in cells and human airway epithelia organoids. Polymerase inhibition within in vitro RNA-dependent RNA polymerase assays established for RSV and SARS-CoV-2 revealed transcriptional stalling after incorporation. Once-daily oral treatment was highly efficacious at 5 milligrams per kilogram (mg/kg) in RSV-infected mice or 20 mg/kg in ferrets infected with different SARS-CoV-2 variants of concern, initiated 24 or 12 hours after infection, respectively. These properties define 4′-FIU as a broad-spectrum candidate for the treatment of RSV, SARS-CoV-2, and related RNA virus infections.

T he COVID-19 experience has highlighted the need for orally bioavailable broad-spectrum antivirals that may be quickly deployed against newly emerging viral pathogens. Remdesivir—a direct-acting broad-spectrum antiviral—is still the only small-molecule therapeutic approved for use against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in the United States, but it requires intravenous administration. The ensuing restriction to hospitalized patients compromises its clinical effect as treatment is initiated too late in the infection cycle (1). We have demonstrated the

efficacy of orally available EIDD-2801 (molnupiravir) against influenza viruses in human organoid models and ferrets (2), and subsequent animal and human data showed that the antiviral efficacy of molnupiravir extends to SARS-CoV-2 in vivo (3, 4). Molnupiravir acts by inducing lethal viral mutagenesis after incorporation into viral genomic RNA of influenza viruses (2) and betacoronaviruses (5). The drug was recently approved in the United Kingdom and is currently considered for emergency use authorization against COVID-19 in the United States. However, even with this accelerated development timeline, molnupiravir only became available to patients nearly 2 years into the pandemic. To have a substantial effect on a mounting pandemic, an antiviral must be approved for human use before a new pathogen emerges, making the case for the development of broad-spectrum antivirals.

We have identified respiratory syncytial virus (RSV) disease as a viable primary indication for a candidate broad-spectrum antiviral, on the basis of the unaddressed major health threat imposed by RSV and well-established protocols for clinical trials of anti-RSV therapeutics. RSV infections are responsible for over 58,000 hospitalizations of children <5 years of age in the United States annually, and ~177,000 hospitalizations of adults > the age of 65 (6–9). Despite this major health and economic burden, no therapeutics have been licensed specifically

Fig. 1. 4′-FIU is a potent broad-spectrum antiviral. (**A**) Chemical structure of 4′-FIU. (**B**) Virus yield reduction of RSV clinical isolates 6A8, 16F10, 2-20, and recombinant recRSV-A2line19F-[mKate] [(A) or (B) antigenic subgroup]. (**C**) HEp-2, MDCK, BHK-T7, and BEAS-2B cell lines were assayed for reduction in cell metabolism by 4′-FIU. (**D** and **E**) recRSV-A2line19F-[FireSMASh] dose response inhibition and cytotoxicity assay with human airway epithelial (HAE) cells (D) from two donors in the presence of indicated 4′-FIU concentrations (E). (**F**) Dose-response inhibition of a panel of recombinant mononegaviruses by 4′-FIU. (**G**) Dose-response inhibition of recSARS-CoV-2-[Nluc] and virus yield reduction of alpha, gamma, and delta VoC isolates by 4′-FIU. (**H**) Dose-response inhibition of transiently expressed polymerase complexes from mononegaviruses MeV, RSV, NiV, or HPIV-3 by 4′-FIU (**I**) recRSV-A2line19F-[FireSMASh]-infected cells were treated with 10 μM of 4′-FIU and serial dilutions of exogenous nucleotides in extracellular media. Viral activity was determined by reporter activity. Symbols represent independent repeats [(B), (E), (G), (H), and (I)] or mean with standard deviation [(C) and (F)], and lines represent means. $n \geq 3$, EC_{50}s, and CC_{50}s are reported in tables S1 and S2, and all source data are provided in data S2.
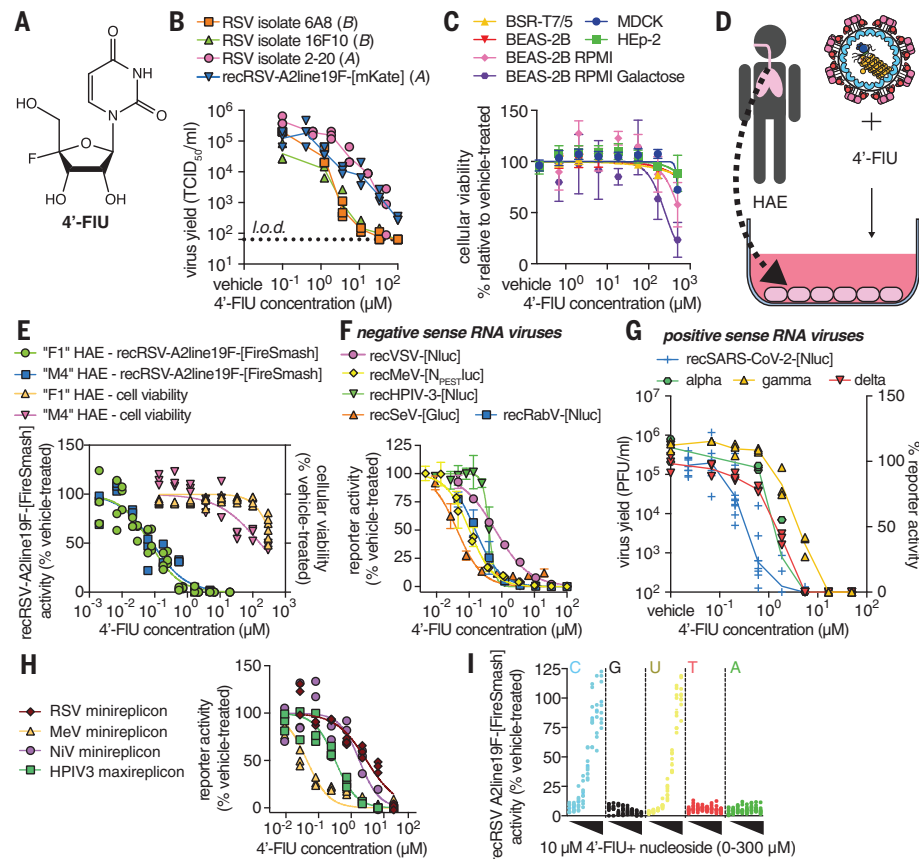
for treatment of RSV disease (*10*). Anti-RSV drug discovery efforts have increasingly focused on inhibiting the viral RNA-dependent RNA polymerase (RdRP) complex (*11*). The core polymerase machinery comprises the large (L) polymerase protein, its obligatory cofactor [the phosphoprotein (P)], and the encapsidated negative-sense RNA genome (*11*). Allosteric inhibitors of RSV L have potent activity as seen, for instance, with the experimental drug candidates AVG-233 (*12*) and inhaled PC786 (*13*).

In search of a drug that is active against RSV and SARS-CoV-2, is orally available, and acts through a distinct mechanism of activity (MOA) from molnupiravir, we explored 4′-fluorine substitutions in a series of analogs of the molnupiravir parent molecule N^4-hydroxycytidine (NHC) (*14*). The focus on 4′-fluorine ribose substitutions was motivated by the small atomic radius and strong stereoelectronic effect of fluorine that can influence backbone conformation flexibility, which may lead to improved selectivity indices (SIs), increased lipophilicity, and greater metabolic stability (*15*). A synthetic intermediate in the approach to 4′-fluoro-N^4-hydroxycytidine (compound 5 in fig. S1) was deprotected to provide 4′-FIU (Fig. 1A), which emerged as a broadly active antiviral when biotested.

4′-FIU is a broad-spectrum mononegavirus inhibitor with high SI

Following the approach of using RSV disease as a primary indication to advance a new candidate broad-spectrum antiviral, we first assessed activity of 4′-FlU against a recombinant RSV A2-line19F (recRSV A2-L19F) (*16*) and clinical RSV isolates on immortalized HEp-2 cells. The compound showed potent dose-dependent activity against all RSV strains tested, returning half-maximal effective concentrations (EC$_{50}$) values) ranging from 0.61 to 1.2 μM (Fig. 1B and table S1). This cell culture potency was on par with the previously reported anti-RSV activity of NHC (fig. S2). Global metabolic activity of established human and animal cell lines (HEp-2, MDCK, BHK-T7, and BEAS-2B) remained unaltered after they were exposed to 4′-FlU (up to 500 μM), indicating that the antiviral effect is a result of cytotoxicity (Fig. 1C and table S2). When glucose was replaced with galactose as a carbohydrate source to link cell metabolic activity strictly to mitochondrial oxidation (*17*), we determined a half-maximal

[1]Center for Translational Antiviral Research, Georgia State University, Atlanta, GA 30303, USA. [2]Texas Biomedical Research Institute, San Antonio, TX 78227, USA. [3]Emory Institute for Drug Development, Emory University, Atlanta, GA 30322, USA. [4]Drug Innovation Ventures at Emory (DRIVE), Atlanta, GA 30322, USA. [5]Department of Pharmacology, Emory University School of Medicine, Atlanta, GA 30322, USA. [6]Department of Pediatrics, Emory University School of Medicine, Atlanta, GA 30322, USA.
*Corresponding author. Email: rplemper@gsu.edu

Fig. 2. 4′-FIU induces a delayed stalling of RSV and SARS-CoV-2 RdRP. (A) SDS-PAGE with Coomassie blue staining of recombinant RSV RdRP complexes (L and P proteins). (B) Schematics of the primer extension assay. (C) Urea-polyacrylamide gel electrophoresis (PAGE) fractionation of RNA transcripts produced through primer extension by the RSV RdRP in the presence of the indicated nucleotides (n = 3). (D) Kinetic analysis of autoradiographs from (C). Nonlinear regression with the Michaelis-Menten model. K_m and V_{max} with 95% confidence intervals (CIs) and goodness of fit (r^2) are indicated. (E to G) Urea-PAGE fractionation of RNA transcripts produced by RSV RdRP in the presence of the indicated templates and nucleotides. "Remdesivir" denotes the addition of the remdesivir active metabolite GS-443902, a well-characterized "delayed chain terminator". 4′-FIU-TP bands in (F) to (G) were normalized to the corresponding band after UTP incorporation. Bars represent mean and error bars represent standard deviation (n = 3). (H) Purified recombinant SARS-CoV-2 RdRP complexes (nsp7, 8, and 12 proteins) "nsp12 SNN" denotes a catalytically inactive mutant. (I to K) Urea-PAGE fractionation of RNA transcripts produced by SARS-CoV-2 RdRP in the presence of the indicated templates and nucleotides. Stars denote cellular contaminants. Uncropped autoradiograph replicates are provided in data S1.

cytotoxic concentration (CC$_{50}$) of 4′-FlU of 250 μM (Fig. 1C and table S2).

When tested on disease-relevant primary human airway epithelial cells (HAEs) derived from two different donors (Fig. 1D), 4′-FlU showed a ≥17-fold increase in anti-RSV potency relative to that on HEp-2 cells; however, the low cytotoxicity levels remained unchanged (CC$_{50}$ 169 μM) (Fig. 1E), resulting in a high SI (SI = EC$_{50}$/CC$_{50}$) of ≥1877. Consistent with these findings, quantitative immunocytochemistry on HAE cells confirmed that 4′-FlU reduced steady-state levels of nuclear- (SDH-A; IC$_{50}$ 272.8 μM) and mitochondrial- (COX-I; IC$_{50}$ 146.8 μM) encoded proteins only at high concentrations (fig. S3).

We next explored the 4′-FlU indication spectrum. We assessed a panel of negative-sense RNA viruses of the paramyxovirus and rhabdovirus families, including measles virus (MeV), human parainfluenza virus type 3 (HPIV3), Sendai virus (SeV), vesicular stomatitis virus (VSV), and rabies virus (RabV). Like RSV, these viruses belong to the mononegavirus order, and we found that 4′-FlU demonstrated submicromolar active concentrations (Fig. 1F and table S1). Testing a representative of phylogenetically distant positive-sense RNA viruses, the betacoronavirus SARS-CoV-2 was also sensitive to 4′-FlU, with EC$_{50}$ values ranging from 0.2 to 0.6 μM against isolates of different lineages (Fig. 1G and table S1).

At initial mechanistic characterization, 4′-FlU inhibited RSV and paramyxovirus RdRP complex activity in cell–based minireplicon systems (Fig. 1H and table S1). The RdRP activity of Nipah virus (NiV)—a highly pathogenic zoonotic paramyxovirus with pandemic potential (18)—was also efficiently inhibited by 4′-FlU in a NiV minireplicon reporter assay. The antiviral effect of 4′-FlU was dose-dependently reversed by addition of an excess of exogenous pyrimidines (cytidine and uridine)—but not purines—to the cultured cells, which is consistent with competitive inhibition of RdRP activity (2, 19) (Fig. 1I).

Incorporation of 4′-FlU by RSV and SARS-CoV-2 RdRP causes sequence-modulated transcriptional stalling

To characterize the molecular MOA of 4′-FlU, we purified recombinant RSV L and P proteins expressed in insect cells (Fig. 2A) and determined performance of the bioactive 5′-triphosphate form of 4′-FlU (4′-FlU-TP) within in vitro primer extension assays (20) (Fig. 2B). In the presence of radio-labeled adenosine triphosphate (ATP) and an increasing amount of uridine triphosphate (UTP), RSV RdRP complexes elongated the primer until reaching a G in third position on the template strand, and continued further upon addition of CTP (Fig. 2C) (fig. S4 and data S1). Replacing UTP with 4′-FlU-TP resulted in efficient primer exten-
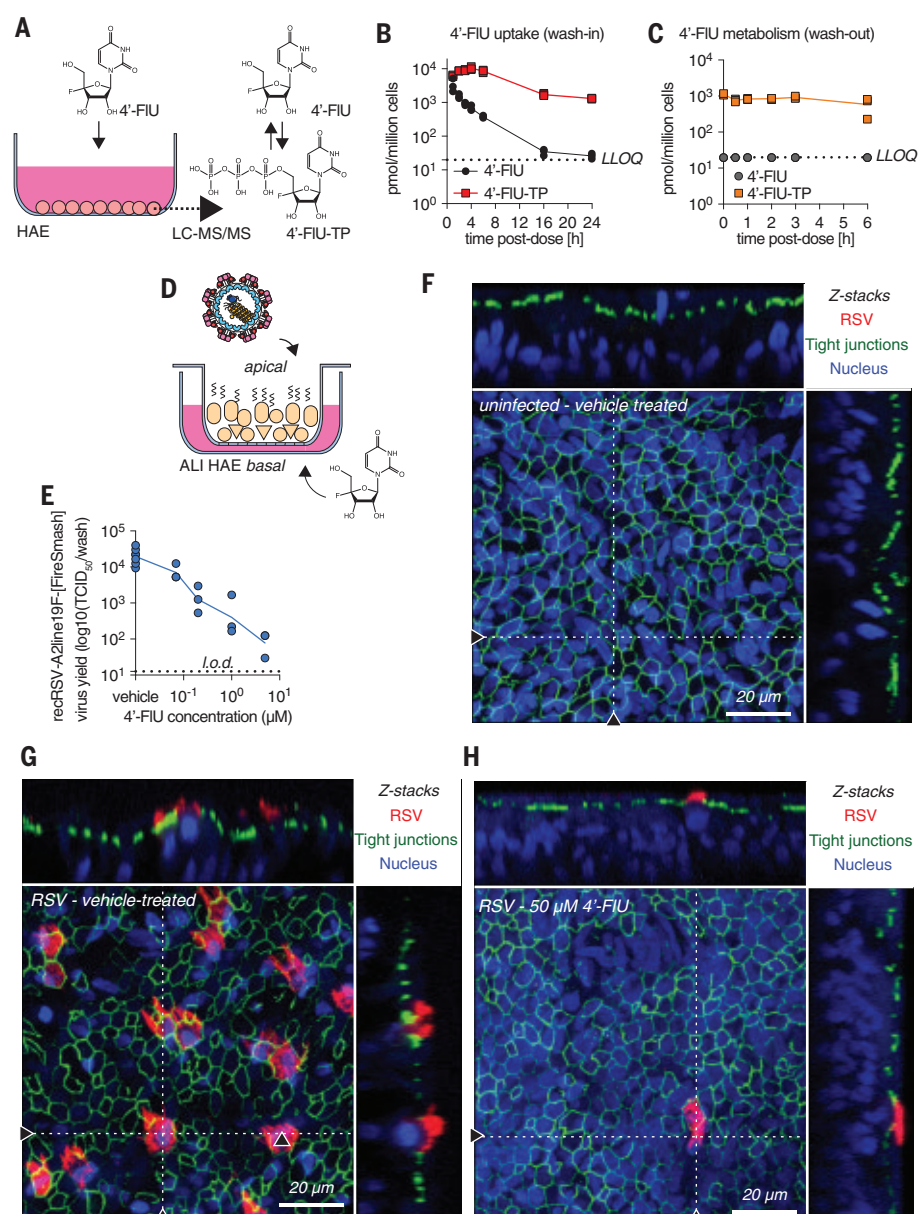
Fig. 3. 4′-FlU is efficiently anabolized in HAE cells and is efficacious in human airway epithelium organoids. (**A** to **C**) 4′-FlU cellular uptake and metabolism in "F1" HAE cells quantified by mass spectrometry (A). Intracellular concentration of 4′-FlU(-TP) after exposure to 20 μM 4′-FlU for 0, 1, 2, 3, 4, 6, 16, and 24 hours (B), or 24-hour incubation followed by removal of the compound for 0, 0.5, 1, 2, 3, and 6 hours before quantification (C) (n = 3). The low limit of quantitation (LLOQ) for 4′-FlU (19.83 pmol/10^6 cells) is indicated by the dashed line. (**D**) HAE cells were matured at the air-liquid interface (ALI). (**E**) Virus yield reduction of recRSV-A2line19F-[FireSMASh] was shed from the apical side in ALI HAE after incubation with serial dilutions of 4′-FlU on the basal side (n = 3). (**F** to **H**) Confocal microscopy of ALI HAE cells infected with recRSV-A2line19F-[FireSMASh], at 5 days after infection. RSV-infected cells, tight junctions, and nuclei were stained with anti-RSV, anti-ZO-1, and Hoechst 34580. z-stacks of 30 1-μm slices with 63 × oil objective. Dotted lines, x-z and y-z stacks; scale bar, 20 μm. In all panels, symbols represent independent biological repeats and lines represent means.

sion up to the third nucleotide, confirming that RSV RdRP recognizes and incorporates 4′-FlU in place of UTP (Fig. 2C). Incorporation kinetics (21) showed only a moderate reduction in substrate affinity for 4′-FlU-TP compared with UTP (Fig. 2D). Further addition

of CTP to the reaction mix resulted in limited elongation rather than the expected full-length product, which suggested delayed polymerase stalling by incorporated 4′-FlU (fig. S4 and data S1). Direct side-by-side comparison with GS-443902—the active metabolite of remdesivir

and a "delayed polymerase stalling" inhibitor well characterized for SARS-CoV-2—along with RSV and other RNA viruses (21, 22), corroborated this antiviral effect of 4′-FlU-TP (Fig. 2E and data S1).

When a modified template coded for incorporation of only a single UTP (Fig. 2F and data S1), primers elongated preferentially to position $i + 3$ after 4′-FlU-TP, whereas the efficiency of full elongation was strongly reduced compared with extension in the presence of UTP. However, repositioning the incorporation site further downstream in the template triggered immediate polymerase stalling at position i (fig. S5), indicating template sequence dependence of the inhibitory effect. Transcription stalling at i or $i + 3$ was also observed after multiple 4′-FlU incorporations: An AxAxxx template (Fig. 2G) and direct tandem incorporations through an AAxxAx template (fig. S5) caused stalling at position i,

whereas increasing spacer length between the incorporated uridines shifted preferential stalling to $i+3$ (fig. S5). This variable delayed polymerase-stalling event within one to four nucleotides of the incorporation site was equally prominent when we examined de novo initiation of RNA synthesis at the promoter with a synthetic native RSV promoter sequence rather than extension of primer-template pairs (fig. S6).

Purification of a core SARS-CoV-2 polymerase complex [nonstructural proteins (nsp) 7, 8, and 12] from bacterial cell lysates (23, 24) (Fig. 2H) and assessment of RdRP bioactivity in equivalent primer-extension in vitro polymerase assays (Fig. 2I) again demonstrated incorporation of 4′-FlU-TP in place of UTP by the coronavirus RdRP (Fig. 2J), but there was no sign of immediate polymerase stalling. However, SARS-CoV-2 polymerase stalling was triggered by multiple incorporations of 4′-FlU-TP, and was particularly prominent when a sec-

ond incorporation of 4′-FlU-TP occurred at the $i+4$ position (Fig. 2K and fig. S7). Primer extension was blocked when the nsp12 subunit was omitted or an nsp12 variant carrying mutations in the catalytic site was used, confirming specificity of the reaction (fig. S7 and data S1).

4′-FlU is rapidly anabolized, metabolically stable, and potently antiviral in disease-relevant well-differentiated HAE cultures

Quantitation of 4′-FlU and its anabolites in primary HAE cells (Fig. 3A) demonstrated rapid intracellular accumulation of 4′-FlU, reaching a level of 3.42 nmol/million cells in the first hour of exposure (Fig. 3B). Anabolism to bioactive 4′-FlU-TP was efficient, resulting in concentrations of 10.38 nmol per million cells at peak (4 hours after start of exposure) and 1.31 nmol per million cells at plateau (24 hours). The anabolite was metabolically stable, remaining present in sustained concentrations of
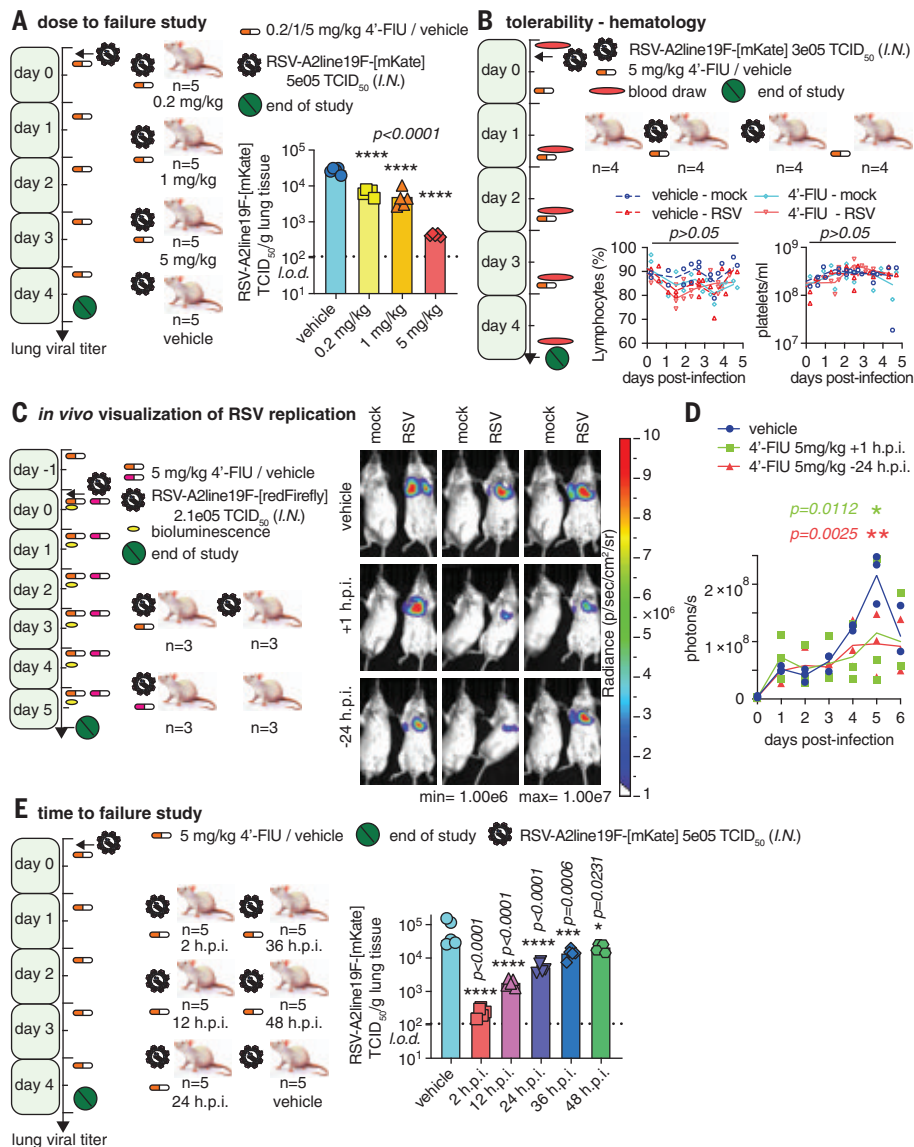
Fig. 4. Therapeutic oral efficacy of 4′-FlU in the RSV mouse model. (**A**) Balb/cJ mice were inoculated with recRSV-A2line19F-[mKate] and treated as indicated. At 4.5 days after infection, viral lung titers were determined with $TCID_{50}$ titration (n = 5). (**B**) Balb/cJ mice were inoculated with recRSV-A2line19F-[mKate] or mock-infected, and treated as indicated. Blood samples were collected before infection and at 1.5, 2.5, 3.5, and 4.5 days after infection; lymphocyte proportions with platelets/ml are represented over time (n = 4). (**C**) Balb/cJ mice were inoculated with recRSV-A2line19F-[redFirefly] and treated as indicated. In vivo luciferase activity was measured daily. (**D**) Total photon flux from mice lungs from (C) over time (n = 3). (**E**) Balb/cJ mice were inoculated with recRSV-A2line19F-[mKate] and treated as indicated. At 4.5 days after infection, viral lung titers were determined with $TCID_{50}$ titration (n = 5). In all panels, symbols represent individual values, and bars or lines represent means. One-way ordinary analysis of variance (ANOVA) with Tukey's post hoc multiple comparisons (B) and (I) or two-way ANOVA with Dunnett's post hoc multiple comparison (C) and (G). h.p.i., hours post-infection.

~1 nmol/million cells over a 6-hour monitoring period, corresponding to an extrapolated half-life of 9.7 hours (Fig. 3C).

To explore efficacy in a disease-relevant human tissue model, we cultured the HAEs at the air-liquid interface, inducing the formation of a well-differentiated three-dimensional (3D) airway epithelium that included ciliated and mucus-producing cells (*25*) (Fig. 3D). Adding 4′-FlU to the basolateral chamber of the transwells after apical infection of the epithelium with RSV potently reduced apical virus shedding with an EC_{50} of 55 nM (Fig. 3E). Overall titer reduction spanned nearly four orders of magnitude, ranging from 3.86×10^4 median tissue culture infectious dose ($TCID_{50}$) in control cells to 78.18 $TCID_{50}$ at 5 µM basolateral 4′-FlU, approaching the level of detection.

Confocal microscopy validated formation of a pseudostratified organization of the epithelium with tight junctions in the airway epithelium tissue model (Fig. 3F), visualized efficient RSV replication in vehicle-treated tissue models (Fig. 3G), and confirmed near-sterilizing antiviral efficacy in the presence of 50 µM basolateral 4′-FlU (Fig. 3H and figs. S8 and S9). Under the latter conditions, positive staining for RSV antigens was rarely detected.

4′-FlU is orally efficacious in a therapeutic dosing regimen in a small-animal model of RSV infection

To test 4′-FlU efficacy in vivo, we used the mouse model of RSV infection (supplementary text), challenging animals with recRSV-A2-L19F, which efficiently replicates in mice (*16*). In a dose-to-failure study, we infected BALB/cJ mice intranasally and initiated once-daily oral treatment 2 hours after infection at 0.2, 1, or 5 mg 4′-FlU per kilogram of body weight. Treatment at all dose levels resulted in a statistically significant reduction in lung virus load compared with vehicle-treated animals (Fig. 4A). The antiviral effect was dose dependent and approached nearly two orders of magnitude at the 5 mg/kg dose. Consistent with high metabolic stability in HAEs, a twice-daily dosing regimen did not significantly enhance efficacy (fig. S10). Because animal appearance, body weight, temperature (fig. S11), and relative lymphocyte and platelet counts (Fig. 4B and fig. S12) were unchanged in the 5 mg/kg group compared with vehicle-treated animals, we selected this dose for further studies.

For a longitudinal assessment of therapeutic benefit, we used an in vivo imaging system (IVIS) with a red-shifted luciferase (*26*) expressing an RSV reporter virus generated for this study. This assay allows for a noninvasive spatial appreciation of intrahost viral dissemination. Daily imaging (Fig. 4C and fig. S13) revealed considerable reduction of biolumi-

nescence intensity in lungs of 4′-FlU-treated animals at 5 days after infection, corresponding to peak viral replication, independent of whether treatment was initiated 24 hours before or 1 hour after infection (Fig. 4D). This IVIS profile is consistent with reduced viral replication and ameliorated viral pneumonia in treated animals.

To probe the therapeutic window of 4′-FlU, we initiated treatment at 2, 12, 24, 36, and 48 hours after infection. All treatment groups showed a statistically significant reduction of lung virus burden compared with vehicle-treated animals, but effect size was dependent on the time of treatment initiation (Fig. 4E and fig. S14). On the basis of our experience with
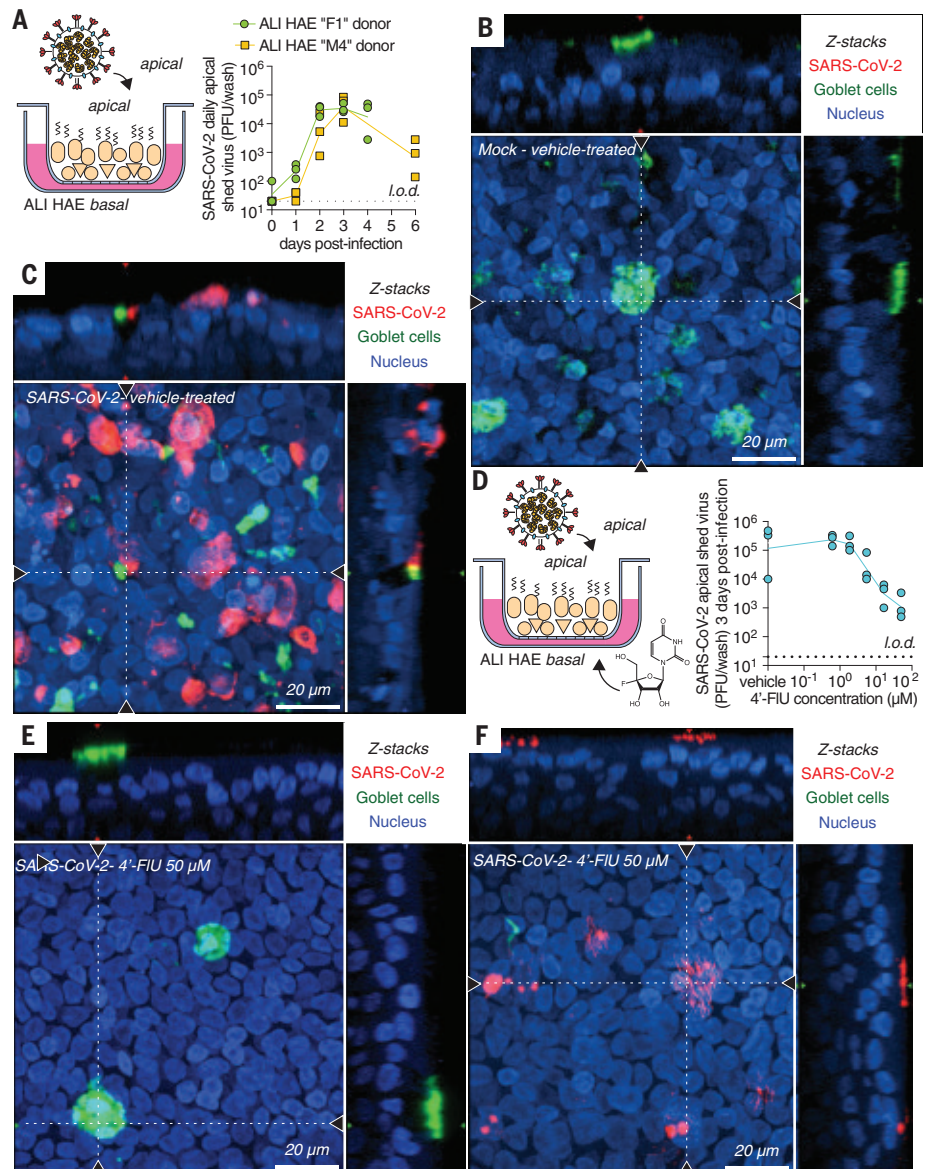
Fig. 5. Efficacy of 4′-FlU against SARS-CoV-2 replication in HAE organoids. (**A**) Multicycle growth curve of SARS-CoV-2 WA1 isolate on ALI HAE from two donors. Shed virus was harvested daily and titered by plaque assay (*n* = 3). (**B** and **C**) Confocal microscopy of ALI HAE cells from "F1" donor mock-infected (B) or infected (C) with SARS-CoV-2 WA1 isolate, 3 days after infection. SARS-CoV-2 infected cells, goblet cells, and nuclei were stained with anti-SARS-CoV-2 N immunostaining, anti-MUC5AC immunostaining, and Hoechst 34580, pseudocolored in red, green, and blue, respectively. z-stacks of 35-µm slices (1 µm thick) with 63× objective with oil immersion. Dotted lines represent the location of x-z and y-z stacks; scale bar, 20 µm. In all panels, symbols represent independent biological repeats and lines represent means. (**D**) Virus yield reduction of SARS-CoV-2 WA1 clinical isolate shed from the apical side in ALI HAE after incubation with serial dilutions of 4′-FlU on the basal side (*n* = 3). (**E** and **F**) Confocal microscopy of ALI HAE cells infected with SARS-CoV-2 WA1 isolate and treated with 50 µM 4′-FlU 3 days after infection. Rare ciliated cells positive for N are represented in (F).
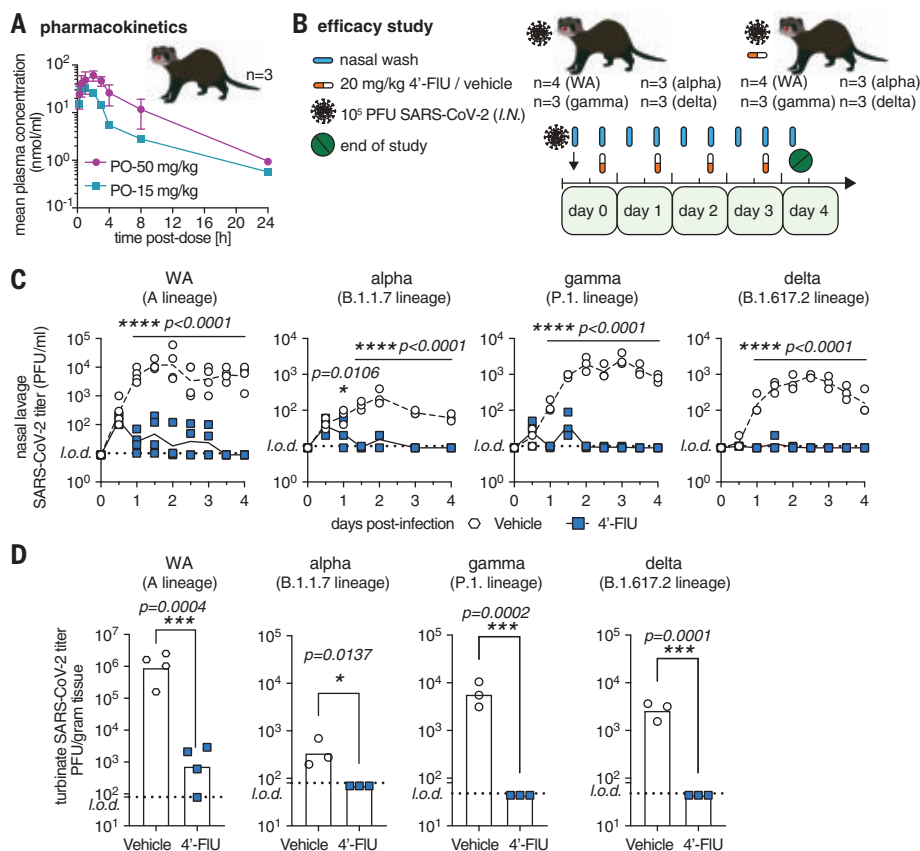
Fig. 6. Therapeutic oral efficacy of 4′-FlU against different SARS-CoV-2 isolates in ferrets. (**A**) Single oral dose (15 or 50 mg/kg bodyweight) pharmacokinetic properties of 4′-FlU in ferret plasma (*n* = 3). (**B**) Ferrets were inoculated with SARS-CoV-2 WA1 or VoC alpha, gamma, or delta, and treated as indicated. (**C**) Nasal lavages were performed twice daily and viral titers were determined by plaque assay [*n* = 4 (WA1) or *n* = 3 (alpha, gamma, delta)]. (**D**) Viral titers in nasal turbinates 4 days after infection. In all panels, symbols represent individual independent biological repeats and lines show mean values. Two-way ANOVA with Sidak's post hoc multiple comparison (C) and unpaired *t* test (D).

therapeutic intervention in related respiratory RNA viruses that cause lethal disease (*25*), we required a reduction in lung virus load of at least one order of magnitude. With this constraint, the therapeutic window of 4′-FlU was extended to 24 hours after infection in mice.

4′-FlU is effective against SARS-CoV-2 in HAE and the ferret model

To test activity against SARS-CoV-2 in the human airway organoids, we first confirmed that the WA1 isolate replicated efficiently in the HAEs of all donors tested (Fig. 5, A to C, and fig. S15). Treatment of infected organoids with basolateral 4′-FlU dose-dependently reduced apical virus shedding, albeit with a limited maximal effect size of ~ two orders of magnitude at 50 μM (Fig. 5D). Confocal microscopy revealed that the epithelium was largely devoid of SARS-CoV-2 nucleocapsid proteins under these conditions (Fig. 5E), with only sporadic staining detectable in a small subset of ciliated cells (Fig. 5F and fig. S15).

To probe for a corresponding antiviral effect in vivo, we determined the efficacy of oral 4′-FlU against an early pandemic isolate (WA1) and VoC alpha, gamma, and delta in the ferret model (*27*), which recapitulates hallmarks of uncomplicated human infection (*3*). For dose level selection in ferrets, we determined single oral dose ferret pharmacokinetic (PK) profiles of 4′-FlU. When administered at 15 or 50 mg/ kg, peak plasma concentrations (C$_{max}$) of 4′-FlU reached 34.8 and 63.3 μM, respectively, and overall exposure was 154 ± 27.6 and 413.1 ± 78.1 hours×nmol/ml, respectively, revealing good oral dose-proportionality (Fig. 6A and table S3). On the basis of this PK performance, we selected once-daily dosing at 20 mg/kg body weight for efficacy tests (Fig. 6B).

Intranasal infection of ferrets with 1×10^5 PFU of each isolate resulted in rapid viral shedding into the upper respiratory tract, which plateaued in vehicle-treated animals 48 to 60 hours after infection (Fig. 6C). Therapeutic treatment with 4′-FlU initiated 12 hours

after infection reduced virus burden in nasal lavages by approximately three orders of magnitude (WA1) to <50 PFU/ml within 12 hours of treatment onset. All three VoC were highly sensitive to 4′-FlU, remaining below the level of detection 36 to 48 hours after onset of oral treatment. Viral titers in nasal turbinate tissue extracted 4 days after infection (Fig. 6D) and associated viral RNA copy numbers (fig. S16) correlated with this reduction in shed virus load. Shedding of infectious particles ceased completely in all animals after 2.5 days of treatment (3 days post-infection).

Conclusions

This study identifies and characterizes the ribonucleoside analog 4′-FlU, which potently inhibits pathogens of different clinically relevant negative and positive-sense RNA virus families. The compound causes delayed stalling of RSV and SARS-CoV-2 polymerases within in vitro RdRP assays, reminiscent of the antiviral effect of remdesivir (*28, 29*). However, 4′-FlU can also trigger immediate RdRP stalling depending on sequence context, suggesting steric hindrance of polymerase advance or of accommodating the next incoming nucleotide as the underlying MOA. We cannot exclude that additional effects further enhance the antiviral effect in cellula as proposed for other nucleoside analogs (*30*). Slightly lower sensitivity of SARS-CoV-2 to 4′-FlU compared with RSV could be a result of the exonuclease activity of the coronavirus polymerase, which can eliminate ribonucleoside analogs (*31, 32*). Alternatively, coronavirus RdRP may have a greater capacity to tolerate the compound, because SARS-CoV-2 RdRP showed a higher tendency than RSV polymerase to advance after 4′-FlU-TP incorporation in the RdRP assays, which do not contain exonuclease functionality.

Once-daily oral administration to mice and ferrets significantly reduced the burden of RSV and SARS-CoV-2, respectively, when treatment was initiated up to 24 (RSV) or 12 (SARS-CoV-2) hours after infection. Because RSV (*33*) and SARS-CoV-2 (*34*) host invasion is slower in humans, these data outline a viable therapeutic window for human treatment. Equally potent activity against SARS-CoV-2 VoC alpha, gamma, and delta demonstrated broad anticoronavirus efficacy of 4′-FlU, building confidence that the compound will remain active against future VoC that may be increasingly less responsive to spike-targeting vaccines or antibody therapeutics. Formal tolerability studies are pending, but 4′-FlU was well tolerated by the human organoid models and efficacious in murids and mustelids. Blood analysis of treated mice uncovered no antiproliferative effect of 4′-FlU on the hematopoietic system. These results establish 4′-FlU as a broad-spectrum orally efficacious inhibitor

of major RNA viruses, making it a promising therapeutic option for RSV disease and COVID-19, and a much-needed contributor to improvement of pandemic preparedness.

REFERENCES AND NOTES

1. J. H. Beigel *et al.*, *N. Engl. J. Med.* **383**, 1813–1826 (2020).
2. M. Toots *et al.*, *Sci. Transl. Med.* **11**, eaax5866 (2019).
3. R. M. Cox, J. D. Wolf, R. K. Plemper, *Nat. Microbiol.* **6**, 11–18 (2021).
4. G. R. Painter *et al.*, *Antiviral Res.* **171**, 104597 (2019).
5. T. P. Sheahan *et al.*, *Sci. Transl. Med.* **12**, eabb5883 (2020).
6. A. R. Falsey, P. A. Hennessey, M. A. Formica, C. Cox, E. E. Walsh, *N. Engl. J. Med.* **352**, 1749–1759 (2005).
7. B. Rha *et al.*, *Pediatrics* **146**, e20193611 (2020).
8. T. Shi *et al.*, *J. Infect. Dis.* **222** (suppl. 7), S577–S583 (2020).
9. T. Shi *et al.*, *Lancet* **390**, 946–958 (2017).
10. G. S. Cockerill, J. A. D. Good, N. Mathews, *J. Med. Chem.* **62**, 3206–3227 (2019).
11. R. Fearns, R. K. Plemper, *Virus Res.* **234**, 87–102 (2017).
12. R. M. Cox *et al.*, *J. Biol. Chem.* **293**, 16761–16777 (2018).
13. J. DeVincenzo *et al.*, *J. Infect. Dis.* jiaa716 (2020).
14. J. J. Yoon *et al.*, *Antimicrob. Agents Chemother.* **62**, e00766-18 (2018).
15. P. Richardson, *Expert Opin. Drug Discov.* **16**, 1261–1286 (2021).
16. M. L. Moore *et al..*, *J. Virol.* **83**, 4185–4194 (2009).
17. L. D. Marroquin, J. Hynes, J. A. Dykens, J. D. Jamieson, Y. Will, *Toxicol. Sci.* **97**, 539–547 (2007).
18. S. P. Luby, *Antiviral Res.* **100**, 38–43 (2013).
19. Based on the current knowledge of the pyrimidine salvage pathway, it must be assumed that the large excess of extracellular cytidine leads to conversion of CTP to UTP via cytidine deaminase (CDA), competing with 4'-FIU.
20. B. Ludeke, R. Fearns, *Virology* **540**, 66–74 (2020).
21. E. P. Tchesnokov, J. Y. Feng, D. P. Porter, M. Götte, *Viruses* **11**, 326 (2019).
22. C. J. Gordon *et al.*, *J. Biol. Chem.* **295**, 6785–6797 (2020).
23. T. L. Dangerfield, N. Z. Huang, K. A. Johnson, Expression and purification of tag-free SARS-CoV-2 RNA-dependent RNA polymerase in *Escherichia coli*. *STAR Protoc* **2**, 100357 (2021).
24. Our nsp12 preparations contain a mixture of 'GroEL' and 'tf', and nsp12 as described before, the latter representing 19% and 18% of the protein content of the preparations containing catalytically active and inactive nsp12, respectively.
25. M. Toots *et al.*, *Transl. Res.* **218**, 16–28 (2020).
26. S. M. Coleman, A. McGregor, *Future Virol.* **10**, 169–183 (2015).
27. Male ferrets were not included in this study for practical reasons as they are territorial and available ABSL3 space cannot accommodate the number of cages that this study would require to achieve meaningful statistical power.
28. G. Kokic *et al.*, *Nat. Commun.* **12**, 279 (2021).
29. E. P. Tchesnokov *et al.*, *J. Biol. Chem.* **295**, 16156–16165 (2020).
30. J. J. Feld, J. H. Hoofnagle, *Nature* **436**, 967–972 (2005).
31. E. C. Smith, H. Blanc, M. C. Surdel, M. Vignuzzi, M. R. Denison, *PLOS Pathog.* **9**, e1003565 (2013).
32. E. Minskaia *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5108–5113 (2006).
33. C. M. El Saleeby, A. J. Bush, L. M. Harrison, J. A. Aitken, J. P. Devincenzo, *J. Infect. Dis.* **204**, 996–1002 (2011).
34. B. Oberfeld *et al.*, *Cell* **181**, 954–954.e1 (2020).

NEURODEGENERATION

Cryo-EM structures of amyloid-β 42 filaments from human brains

Yang Yang[1]†, Diana Arseni[1]†, Wenjuan Zhang[1]†‡, Melissa Huang[1], Sofia Lövestam[1], Manuel Schweighauser[1], Abhay Kotecha[2], Alexey G. Murzin[1], Sew Y. Peak-Chew[1], Jennifer Macdonald[1], Isabelle Lavenir[1], Holly J. Garringer[3], Ellen Gelpi[4], Kathy L. Newell[3], Gabor G. Kovacs[4,5], Ruben Vidal[3], Bernardino Ghetti[3]*, Benjamin Ryskeldi-Falcon[1]*, Sjors H. W. Scheres[1]*, Michel Goedert[1]*

Filament assembly of amyloid-β peptides ending at residue 42 (Aβ42) is a central event in Alzheimer's disease. Here, we report the cryo–electron microscopy (cryo-EM) structures of Aβ42 filaments from human brains. Two structurally related S-shaped protofilament folds give rise to two types of filaments. Type I filaments were found mostly in the brains of individuals with sporadic Alzheimer's disease, and type II filaments were found in individuals with familial Alzheimer's disease and other conditions. The structures of Aβ42 filaments from the brain differ from those of filaments assembled in vitro. By contrast, in *App*^NL-F knock-in mice, Aβ42 deposits were made of type II filaments. Knowledge of Aβ42 filament structures from human brains may lead to the development of inhibitors of assembly and improved imaging agents.

Alzheimer's disease is defined by the simultaneous presence of two different filamentous amyloid inclusions in the brain: abundant extracellular plaques of amyloid-β (Aβ) and intraneuronal neurofibrillary tangles of tau (*1*). Genetic evidence has indicated that Aβ is key to the pathogenesis of Alzheimer's disease (*2, 3*). Multiplications of the *APP* gene encoding the Aβ precursor protein, as well as mutations in *APP* and in *PSEN1* and *PSEN2*, the presenilin genes, cause familial Alzheimer's disease. Presenilins form part of the γ-secretase complex that is required for the production of Aβ from *APP*. Although variability in γ-secretase cleavage results in Aβ peptides that vary in size, those of 40 (Aβ40) and 42 (Aβ42) amino acids are the most abundant. Mutations associated with familial Alzheimer's disease increase the ratios of Aβ42 to Aβ40 (*4, 5*), the concentration of Aβ42 (*6*), and/or the assembly of Aβ42 into filaments (*7*).

Three major types of Aβ inclusions are typical of the brain in Alzheimer's disease (*8–11*): diffuse and focal deposits in the parenchyma as well as vascular deposits. Diffuse deposits, which contain loosely packed Aβ filaments, are found in several brain regions, including the entorhinal cortex, presubiculum, striatum, brainstem, cerebellum, and subpial area. Focal deposits, in the form of dense core plaques, contain a spherical core of tightly packed Aβ filaments surrounded by more loosely packed filaments. Dense core plaques are found mostly in the hippocampus and cerebral cortex. In advanced cases of Alzheimer's disease, diffuse and focal Aβ deposits are widespread. In ~80% of cases of Alzheimer's disease, Aβ deposits are also found in the walls of blood vessels (cerebral amyloid angiopathy). Cryo–electron
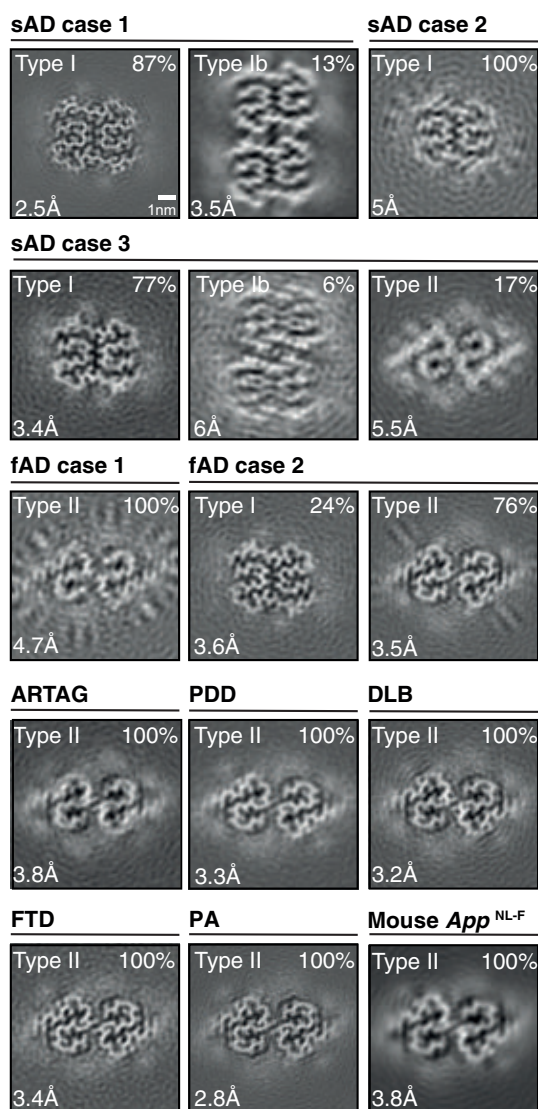
¹Medical Research Council Laboratory of Molecular Biology, Cambridge, UK. ²Thermo Fisher Scientific, Eindhoven, Netherlands. ³Department of Pathology and Laboratory Medicine, Indiana University, Indianapolis, IN, USA. ⁴Institute of Neurology, Medical University, Vienna, Austria. ⁵Tanz Centre and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada.
*Corresponding author. Email: bghetti@iupui.edu (B.G.); bfalcon@mrc-lmb.cam.ac.uk (B.R.-F.); scheres@mrc-lmb.cam.ac.uk (S.H.W.S.); mg@mrc-lmb.cam.ac.uk (M.G.)
†These authors contributed equally to this work. ‡Present address: Medical Research Council Prion Unit, Institute of Prion Diseases, University College London, London, UK.

Fig. 1. Cryo-EM maps of type I, type Ib, and type II Aβ42 filaments from human and mouse brains.
Five cases of Alzheimer's disease [three sporadic (sAD cases 1 to 3) and two familial (fAD case 1, mutation in *APP* encoding V717F, and fAD case 2, mutation in *PSEN1* encoding F105L)], other human diseases (ARTAG, PDD, DLB, FTD caused by a *GRN* mutation, and PA), and homozygous mice of the *App*^NL-F knock-in line. For each map, a sum of the reconstructed densities for several XY-slices, approximating one β rung, is shown. Filament types (type I, type Ib, and type II) are indicated in the top left, and the percentages of a given filament type among Aβ42 filaments in the dataset are shown in the top right. The scale bar shown in the top left panel applies to all panels.

disease, with three sporadic and two familial (a mutation in *APP* encoding V717F and a mutation in *PSEN1* encoding F105L) cases. Five individuals had other conditions—aging-related tau astrogliopathy (ARTAG), Parkinson's disease dementia (PDD), dementia with Lewy bodies (DLB), familial frontotemporal dementia (FTD) caused by a *GRN* mutation, and pathological aging (PA).

Type I Aβ42 filaments from human brains

For individuals with sporadic Alzheimer's disease, we observed a predominance of twisted Aβ filaments, which we named type I filaments (Fig. 1 and Fig. 2, A, B, and D). They are made of two identical S-shaped (a double curve resembling the letter S or its reverse) protofilaments embracing each other with extended arms. The 2.5-Å-resolution map of type I filaments from sporadic Alzheimer's disease case 1 was used to build the atomic model (fig. S2A). The ordered core of each protofilament extends from G9 to A42, with the N-terminal arm consisting of residues 9 to 18 and the S-shaped protofilament consisting of residues 19 to 42. The secondary structure of the protofilaments is composed of five β strands, which are each made of three or more residues. The S-shaped domain folds around two hydrophobic clusters: the N-terminal part around the side chains of F19, F20, V24, and I31 and the C-terminal part around the side chains of A30, I32, M35, V40, and A42 (Fig. 2, B and D, and fig. S2C).

The two protofilaments pack against each other with pseudo-2_1 symmetry (fig. S2E). They form a predominantly hydrophobic interface involving the side chains of L34, V36, V39, and I41 on the outer surface of the S-shaped domain and the side chains of Y10, V12, Q15, and L17 in the N-terminal arm. In sporadic Alzheimer's disease cases 1 and 3, we also observed a minority of type Ib filaments, in which two type I filaments run side by side and are held together by polar interactions, including salt bridges between K16 and E22 (Fig. 1 and fig. S3).

Several additional densities, attributed to ordered solvent molecules, are resolved in the 2.5-Å-resolution cryo-EM map (Fig. 2B and Fig. 3A). One of these, located adjacent to the negatively charged carboxyl groups of E22 and D23 on the filament surface, most likely corresponds to a bound metal ion (Fig. 3, B and C) because the conformations of both acidic residues are restrained, and the binding of metal ions would alleviate the electrostatic repulsion between their negatively charged carboxyl groups. Charged solvent molecules have been proposed to act as cofactors for filament formation by the neutralization of charges on in-register parallel β sheets in amyloids (*25*). By contrast, there are no additional densities associated with an ordered grid of imidazole groups formed on the surface of type I filaments by H13 and H14. Their side chains are

microscopy (cryo-EM) has provided the structures of Aβ40 aggregates from the meninges of Alzheimer's disease brains (*12*). Meningeal deposits have a high Aβ40 and a low Aβ42 content and are morphologically distinct from parenchymal plaques.

Diffuse plaques and the loosely packed material of dense core plaques consist mainly of filamentous Aβ42, whereas plaque cores and blood vessel deposits are made of both Aβ40 and Aβ42. Aβ42 aggregates faster than Aβ40 and is the major species in plaques, despite the proteolytic processing of *APP* generating more soluble Aβ40 (*4, 8, 13*).

Aβ deposition appears to follow spatiotemporal spreading, which suggests that pathology may propagate through seeded aggregation, similarly to what occurs in prion diseases (*14–16*). A prion-like mechanism may also explain the formation of Aβ deposits observed in the cerebral blood vessels in some adults, who received intramuscular injections of contami-

nated human growth hormone preparations as children, and in individuals who were given dura mater grafts or underwent neurosurgery (*17–19*), even though they did not have the symptoms of Alzheimer's disease. Besides Alzheimer's disease, Aβ42 deposits can also be present as a copathology in a number of other conditions, especially as a function of age (*10*). Despite their importance for disease pathogenesis, the structures of Aβ42 filaments from the brain are not known.

We used cryo-EM to determine the structures of Aβ42 filaments extracted from the brains of 10 individuals (Fig. 1, fig. S1, and table S1). When using a sarkosyl extraction method developed for α-synuclein filaments (*20, 21*), we found abundant Aβ42 filaments alongside other amyloids. By contrast, we only observed tau filaments (*22, 23*) when extracting the frontal cortex from individuals with Alzheimer's disease using the standard sarkosyl extraction method (*24*). Five individuals had Alzheimer's

Fig. 2. Structures of type I and type II Aβ42 filaments from the brain. (**A**) Amino acid sequence of Aβ42. Type I filaments (in orange) extend from G9 to A42, and type II filaments (in blue) extend from V12 to A42. Thick connecting lines with arrowheads indicate β strands (β1 to β5 and β1 to β4). (**B** and **C**) Cryo-EM density maps (in transparent gray) and atomic models for type I (B) and type II (C) filaments. Each filament type is made of two identical protofilaments shown in orange (type I) and blue (type II). The density maps are displayed using the zone feature in ChimeraX at a distance of 2 Å. Associated solvent molecules are shown in white, and putative metals are shown in teal (B) and purple (C). (**D** and **E**) Schematics of type I (D) and type II (E) Aβ42 folds. The schematics were produced using atom2svg.py (42). Negatively charged residues are shown in red, positively charged residues in blue, polar residues in green, nonpolar residues in white, sulfur-containing residues in yellow, and glycines in pink. Thick connecting lines with arrowheads indicate β strands. 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.
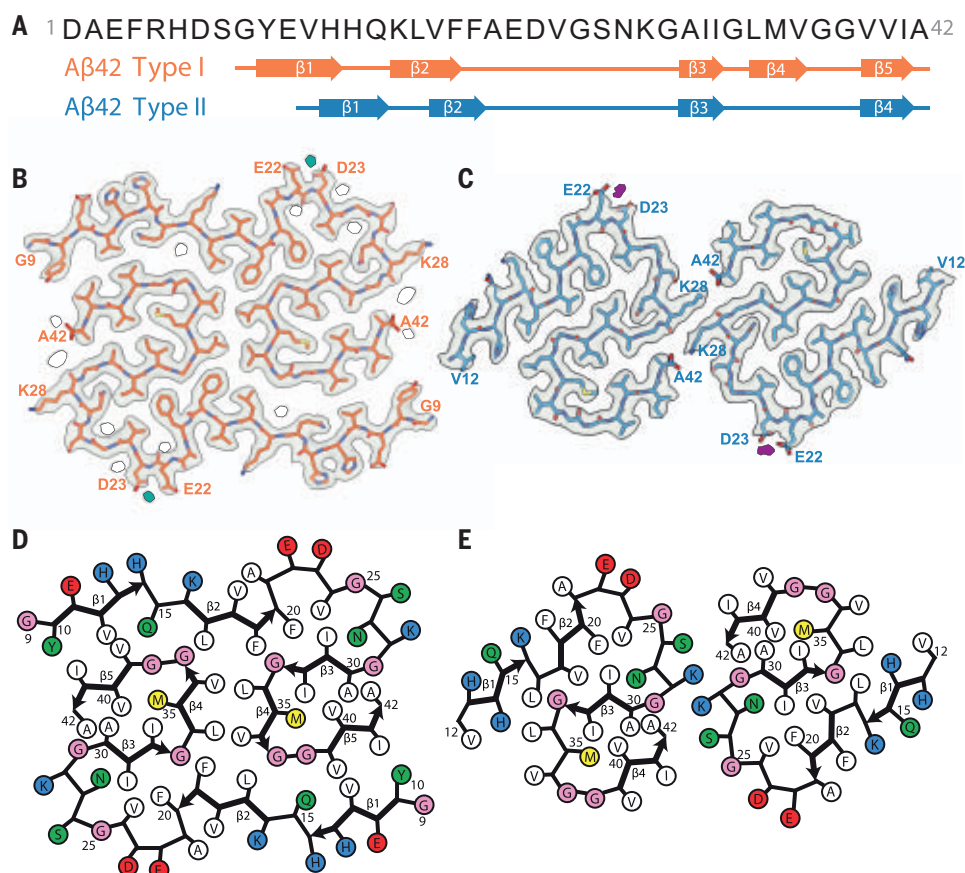
held together by a hydrogen bond, with H13 making a second hydrogen bond with the side chain of E11 in the next Aβ42 molecule.

Type II Aβ42 filaments from human brains

For individuals with familial Alzheimer's disease and other conditions, we observed a major, twisted filament type, distinct from type I, which we named type II (Fig. 1 and Fig. 2, A, C, and E). In case 3 of sporadic Alzheimer's disease, 17% of filaments were type II, whereas in case 2 of familial Alzheimer's disease, 24% of filaments were type I. The atomic model of type II filaments, built using the 2.8-Å-resolution map obtained for the case of PA (fig. S2B), revealed that the ordered core extends from V12 to A42 and is made up of four β strands. Residues 20 to 42 adopt an S-shaped fold similar to that of type I filaments, with the same side chain orientations. Differences between folds are mostly limited to the orientations of a few peptide groups that affect secondary structure assignments. Peptides G25 to S26 and V36 to G37 are flipped by ~180° in the type II fold. The flipped G25 to S26 peptide results in a slight expansion of the N-terminal hydrophobic cluster by accommodating the side chains of L17 and V18 instead of F19, which faces outward in type II filaments (Fig. 3D and fig. S2D). The reorientation of the second peptide

leads to a shift of the C-terminal segment of the type II fold along the helical axis by approximately one Aβ peptide compared with its position in the type I fold (Fig. 3, E and F).

When compared with the type I protofilament interface, that of type II protofilaments is smaller and is formed by the opposite side of the S-shaped fold. Type II protofilaments pack against each other with C2 symmetry (fig. S2F). The protofilament interface is primarily stabilized by electrostatic interactions between the amino group of K28 from one protofilament and the carboxyl group of A42 from the other, and vice versa (Fig. 2C). Unlike type I filaments, hydrophobic residues on the outer surfaces of the S-shaped domains remain exposed, forming nonpolar patches on the surface of type II filaments (fig. S2D). There are fewer additional densities for ordered solvent molecules in the 2.8-Å map of type II filaments than in the 2.5-Å map of type I filaments, but the density for the putative metal ion bound to E22 and D23 is prominent in the equivalent location (Fig. 2, B and C, and Fig. 3, E and F).

Comparison with known structures

Type I and type II filaments have a left-handed twist and are structurally different from Aβ40 aggregates from the meninges of individuals

with Alzheimer's disease, which comprise two identical protofilaments with an unrelated C-shaped fold and a right-handed twist (12). They also differ from the cryo-EM structures of left-handed Aβ40 filaments, which were derived from the cerebral cortex of an individual with Alzheimer's disease by seeded filament growth (26), but share with them a common substructure (Fig. 4A). In the seeded Aβ40 filaments, which comprised two extended protofilaments, residues G25 to G37 adopted virtually the same conformation as that in the middle of the S-shaped fold of type I and type II filaments. Structures of Aβ42 filaments assembled in vitro, obtained by cryo-EM (27) and solid-state nuclear magnetic resonance (NMR) (28–30), each have a single or two identical protofilaments with an S-shaped domain like that of type I and type II filaments (Fig. 4B). In two NMR structures, the interprotofilament packing also resembled that of type I filaments. However, when examined at the single-residue level, none of the Aβ42 filaments assembled in vitro displayed the same side chain orientations and contacts or the same interprotofilament packing as that observed in type I and type II filaments. The structures of in vitro assembled filaments of Aβ40 with the Osaka mutation (deletion of codon 693 in *APP*, corresponding to E22 in Aβ), based on a
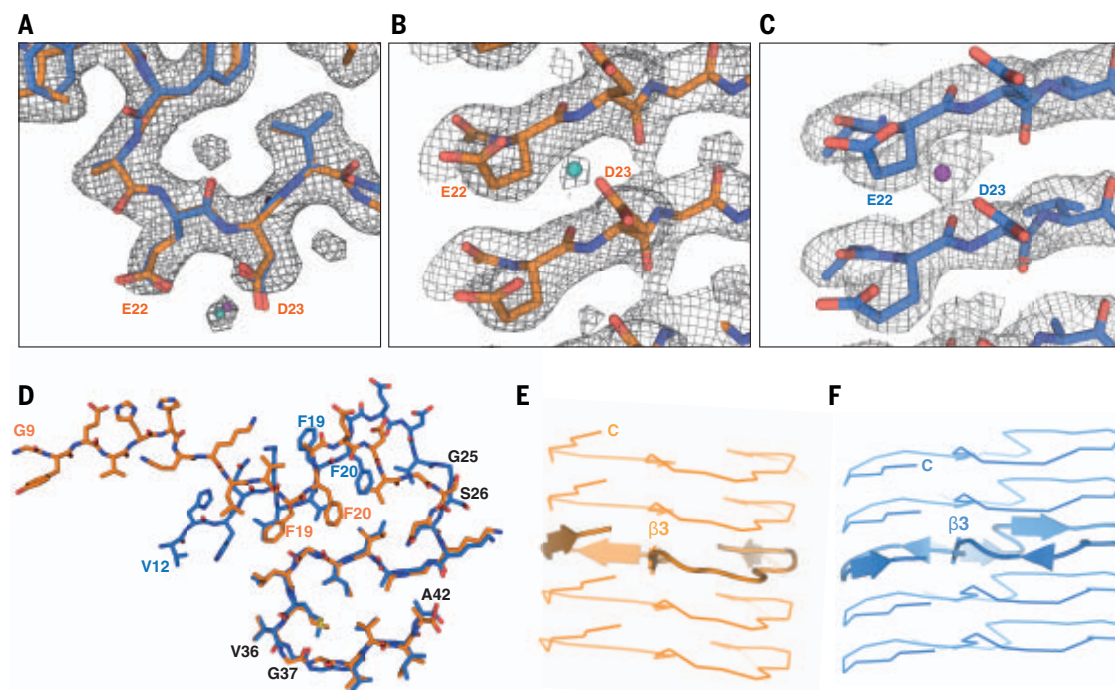
Fig. 3. Protofilament folds and putative metal ion–binding sites of type I and type II Aβ42 filaments. (**A**) Superposition of the structures of F20-V24 arcs overlaid on the corresponding section of the 2.5-Å electron density map of type I filaments. Putative metal ions are shown as teal and purple spheres. (**B** and **C**) Side views of putative metal ion–binding sites in type I (teal) and type II (purple) protofilaments, superimposed on the corresponding density maps. (**D**) Superposition of type I (orange) and type II (blue) protofilaments, based on the central layer of their S-shaped domains. (**E** and **F**) Side views of type I (E) and type II (F) protofilaments along the central β3 strand. The center layer monomers in five successive rungs are shown in cartoon, with β strands shown as arrows.
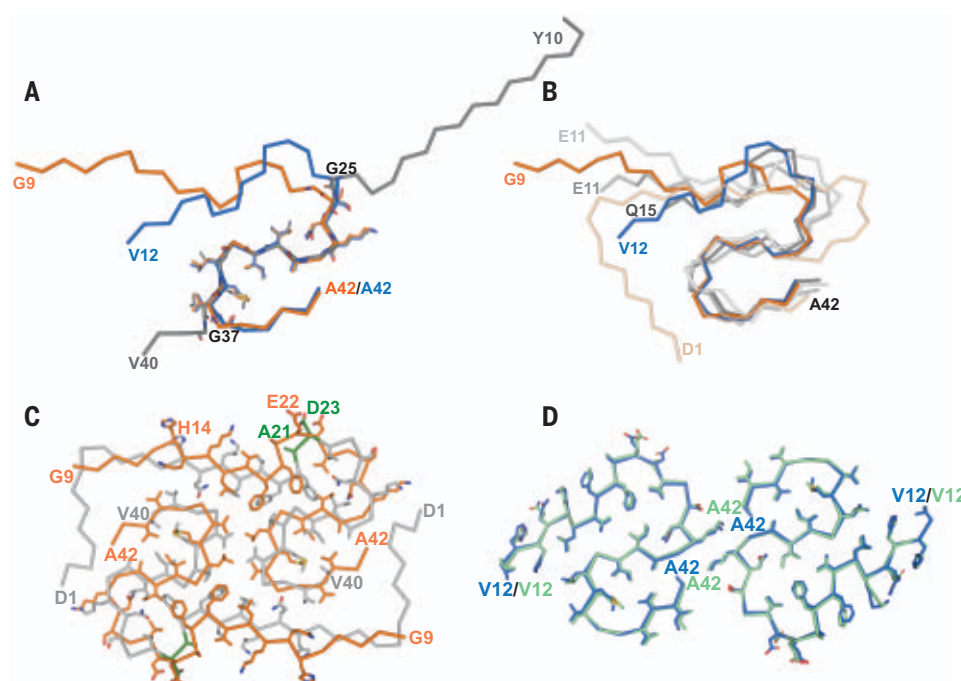
Fig. 4. Comparison of protofilaments and filaments of brain Aβ42 with those of seeded recombinant Aβ40, recombinant Aβ42, and recombinant Aβ40ΔE22. (**A**) Comparison of the cryo-EM structures of human brain type I and type II Aβ42 protofilaments with the cryo-EM structure of seeded recombinant Aβ40 protofilaments. Type I is shown in orange, type II is in blue, and seeded Aβ40 [Protein Data Bank (PDB) ID 6W0O] is in gray. (**B**) Comparison of cryo-EM structures of human brain type I and type II protofilaments with cryo-EM and NMR structures of recombinant protofilaments. PDB IDs and color codes for recombinant Aβ42 filaments are as follows: 5OQV, wheat; 2NAO, dark gray; 2MXU, gray; and 5KK3, light gray. (**C**) Comparison of cryo-EM structures of human brain type I filaments and NMR structures of recombinant Aβ40ΔE22 filaments. Human brain type I is shown in orange, recombinant Aβ40ΔE22 (PDB 2MVX) is in gray, and residues around ΔE22 are shown in green. (**D**) Comparison of cryo-EM structures of Aβ42 filaments from the brains of mouse knock-in line *App*[NL-F] with human brain type II filaments. Mouse brain filaments are in green, and human brain type II filaments are in blue.

large number of unambiguous intra- and intermolecular solid-state NMR distance restraints, are most similar to those of type I Aβ42 filaments (Fig. 4C) (*31*).

Reconstructions of type I and type II filaments show strong densities for residues 41 and 42, which indicates that most molecules correspond to Aβ42. In agreement, immunoblotting (fig. S4) and mass spectrometry (fig. S5) of extracted filaments showed that Aβ42 was the most abundant species in all cases, with variable amounts of Aβ40.

We performed immunohistochemistry on the contralateral sides of the brain regions used for cryo-EM, immunoblotting, and mass spectrometry (figs. S6 and S7). Deposits of Aβ42 were also more numerous than those of Aβ40, with sporadic Alzheimer's disease

cases 1 and 3; familial Alzheimer's disease case 1; and the cases of ARTAG, PDD, DLB, and PA showing almost exclusively Aβ42 deposits. Most deposits of Aβ40 were present in sporadic Alzheimer's disease case 2 and in familial Alzheimer's disease case 2. By immunohistochemistry, Aβ40 deposits were also abundant in FTD. This difference with immunoblotting and mass spectrometry may reflect a hemispheric asymmetry in Aβ deposition. Plaque cores were most numerous in sporadic Alzheimer's disease cases 1 to 3, and blood vessel deposits of Aβ40 were found in sporadic Alzheimer's disease case 2 and in familial Alzheimer's disease case 2. In all cases, diffuse deposits of Aβ were more numerous than focal and blood vessel deposits.

It is possible that low levels of Aβ40, or shorter peptides, may be incorporated in type I and type II filaments. The interprotofilament salt bridge between K28 of one protofilament and A42 of the other—in type II, but not type I, filaments—suggests that it is more likely that hybrid Aβ42-Aβ40 filaments are of type I. This is supported by the structural similarities of type I Aβ42 filaments and filaments of Aβ40 with the Osaka mutation (Fig. 4C) (*31*). We did not find evidence for filaments composed predominantly of Aβ40. However, we cannot rule out that such filaments were present in low amounts or were not extracted as dispersed filaments suitable for cryo-EM reconstruction.

Depending on the filament type, 8 or 11 residues are disordered at the N terminus. The β-site *APP* cleaving enzyme 1 (BACE1) generates the N terminus of Aβ (*32*). BACE1 mainly cleaves at residue 1 of Aβ, but some cleavage at residues 11 or 12 also occurs. Structures of type I and type II filaments from the brain are compatible with the incorporation of shorter peptides. However, by immunoblotting and mass spectrometry (figs. S4 and S5), the bulk of Aβ42 in the extracted filaments was full length. It follows that the N-terminal residues that are not present in type I or type II filament cores form the fuzzy coat of Aβ42 filaments. This is supported by the decoration of type I and type II filaments using antibodies specific for the N-terminal region of Aβ (fig. S8). Tau filaments were unlabeled. The fuzzy coat of Aβ42 filaments therefore composes ~20% of the molecule, with the core making up 80%. By contrast, the fuzzy coat of tau filaments from Alzheimer's disease amounts to >80% (*22, 23*).

In vitro aggregation is essential for studying the molecular mechanisms that underlie amyloid formation. However, available methods for the assembly of recombinant tau and α-synuclein yield filament structures that are different from those of filaments extracted from human brains (*21–23, 33, 34*). The same appears to be true of Aβ42 filaments, which

only partially reproduce the structures from the human brain.

Type II Aβ42 filaments from the *App*NL-F mouse brain

Animal models provide another tool for studying the molecular mechanisms of Alzheimer's disease. *App*NL-F knock-in mice express humanized Aβ and harbor the Swedish double mutation (KM670/671NL) as well as the Beyreuther/Iberian mutation (I716F) in *App* (*35*). They develop abundant deposits of wild-type human Aβ42, neuroinflammation, and memory impairment, without requiring the overexpression of *APP*. To further study the relevance of this mouse model for human disease, we determined the cryo-EM structures of Aβ42 filaments from the brains of 18-month-old homozygous *App*NL-F mice (Fig. 1 and Fig. 4D). They were identical to type II filaments from human brains, providing a mouse experimental system with filament structures like those from human brains. It is possible that cofactors required for the formation of type II filaments are present in the brains of mice from the *App*NL-F knock-in line but missing from in vitro experiments.

Discussion

Type I and type II Aβ42 filaments from the brain are each made of two identical protofilaments, but the protofilaments of type I filaments differ from those of type II. This is unlike tau assembly in the human brain, where a single protofilament gives rise to two or more types of filaments (*36*), and α-synuclein in multiple-system atrophy, where four protofilaments give rise to two different filaments (*21*). Here, type I filaments were limited to cases of sporadic Alzheimer's disease that also had the largest number of plaque cores. A majority of type II filaments was present in cases with abundant diffuse deposits of Aβ and a smaller number of focal plaques with cores. This included cases of familial Alzheimer's disease as well as cases of ARTAG, PDD, DLB, FTD, and PA. Cases of Alzheimer's disease with a majority of type I filaments were older at death than other Alzheimer's and non-Alzheimer's cases, with a majority of type II filaments in the neocortex. There was no correlation between Aβ42 filament type and the *APOE* genotype. The relevance of these differences between type I and type II filaments is not known. Because positron emission tomography compound PiB (Pittsburgh compound B) visualizes Aβ deposits in both sporadic and familial cases of Alzheimer's disease, it probably labels both filament types (*37*).

Like V717F, the mutation in *APP* encoding I716F increases the ratio of Aβ42 to Aβ40 (*4, 38, 39*). This may explain the presence of type II Aβ42 filaments in *App*NL-F mice and in human cases with F717 *APP*. Line *App*NL-F

may therefore be a model for some cases of familial Alzheimer's disease, but not necessarily of sporadic disease.

Differential labeling by luminescent conjugated oligothiophene amyloid ligands has suggested substantial heterogeneity in the molecular architecture of Aβ deposits from the brains of patients with Alzheimer's disease (*40*). Our findings indicate that this heterogeneity is not the result of differences in the structures of Aβ42 filaments. As suggested for Aβ40 (*19, 41*), a single Aβ42 filament type predominated in a given Alzheimer's disease brain. Together with a second filament type, it accounted for the Aβ42 filaments from different cases of Alzheimer's disease, ARTAG, PDD, DLB, FTD, and PA. Knowledge of the structures of Aβ42 filaments from the brain may lead to the development of better in vitro and animal models for these diseases, of inhibitors of Aβ42 assembly, and of imaging agents with increased specificity and sensitivity.

REFERENCES AND NOTES

1. J. M. Long, D. M. Holtzman, *Cell* **179**, 312–339 (2019).
2. J. A. Hardy, G. A. Higgins, *Science* **256**, 184–185 (1992).
3. C. Haass, C. Kaether, G. Thinakaran, S. Sisodia, *Cold Spring Harb. Perspect. Med.* **2**, a006270 (2012).
4. N. Suzuki *et al.*, *Science* **264**, 1336–1340 (1994).
5. D. Scheuner *et al.*, *Nat. Med.* **2**, 864–870 (1996).
6. M. Citron *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11993–11997 (1994).
7. M. Pagnon de la Vega *et al.*, *Sci. Transl. Med.* **13**, eabc6184 (2021).
8. T. Iwatsubo *et al.*, *Neuron* **13**, 45–53 (1994).
9. A. Güntert, H. Döbeli, B. Bohrmann, *Neuroscience* **143**, 461–475 (2006).
10. D. R. Thal, J. Walter, T. C. Saido, M. Fändrich, *Acta Neuropathol.* **129**, 167–182 (2015).
11. J. Attems, K. A. Jellinger, F. Lintner, *Acta Neuropathol.* **110**, 222–231 (2005).
12. M. Kollmer *et al.*, *Nat. Commun.* **10**, 4760 (2019).
13. J. T. Jarrett, E. P. Berger, P. T. Lansbury Jr., *Biochemistry* **32**, 4693–4697 (1993).
14. D. R. Thal, U. Rüb, M. Orantes, H. Braak, *Neurology* **58**, 1791–1800 (2002).
15. M. Meyer-Luehmann *et al.*, *Science* **313**, 1781–1784 (2006).
16. H. H. C. Lau, M. Ingelsson, J. C. Watts, *Acta Neuropathol.* **142**, 17–39 (2021).
17. Z. Jaunmuktane *et al.*, *Nature* **525**, 247–250 (2015).
18. K. Frontzek, M. I. Lutz, A. Aguzzi, G. G. Kovacs, H. Budka, *Swiss Med. Wkly.* **146**, w14287 (2016).
19. Z. Jaunmuktane *et al.*, *Acta Neuropathol.* **135**, 671–679 (2018).
20. A. Tarutani, T. Arai, S. Murayama, S. I. Hisanaga, M. Hasegawa, *Acta Neuropathol. Commun.* **6**, 29 (2018).
21. M. Schweighauser *et al.*, *Nature* **585**, 464–469 (2020).
22. A. W. P. Fitzpatrick *et al.*, *Nature* **547**, 185–190 (2017).
23. B. Falcon *et al.*, *Acta Neuropathol.* **136**, 699–708 (2018).
24. S. G. Greenberg, P. Davies, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5827–5831 (1990).
25. M. R. Sawaya, M. P. Hughes, J. A. Rodriguez, R. Riek, D. S. Eisenberg, *Cell* **184**, 4857–4873 (2021).
26. U. Ghosh, K. R. Thurber, W. M. Yau, R. Tycko, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023089118 (2021).
27. L. Gremer *et al.*, *Science* **358**, 116–119 (2017).
28. Y. Xiao *et al.*, *Nat. Struct. Mol. Biol.* **22**, 499–505 (2015).
29. M. T. Colvin *et al.*, *J. Am. Chem. Soc.* **138**, 9663–9674 (2016).
30. M. A. Wälti *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4976–E4984 (2016).
31. A. K. Schütz *et al.*, *Angew. Chem. Int. Ed.* **54**, 331–335 (2015).
32. R. Vassar *et al.*, *J. Neurochem.* **130**, 4–28 (2014).
33. W. Zhang *et al.*, *eLife* **8**, e43584 (2019).
34. S. Lövestam *et al.*, *FEBS Open Bio* **11**, 999–1013 (2021).

35. T. Saito *et al.*, *Nat. Neurosci.* **17**, 661–663 (2014).
36. Y. Shi *et al.*, *Nature* **598**, 359–363 (2021).
37. C. D. Chen *et al.*, *Acta Neuropathol.* **142**, 689–706 (2021).
38. S. F. Lichtenthaler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1365–1370 (2002).
39. C. Guardia-Laguarta *et al.*, *J. Neuropathol. Exp. Neurol.* **69**, 53–59 (2010).
40. J. Rasmussen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13018–13023 (2017).
41. J. X. Lu *et al.*, *Cell* **154**, 1257–1268 (2013).
42. T. Nakane, atom2svg, version 200213, Zenodo (2020); https://doi.org/10.5281/zenodo.4090925.

MARTIAN GEOLOGY

Organic synthesis associated with serpentinization and carbonation on early Mars

A. Steele[1]*, L. G. Benning[2,3], R. Wirth[2], A. Schreiber[2], T. Araki[4], F. M. McCubbin[5], M. D. Fries[5], L. R. Nittler[1], J. Wang[1], L. J. Hallis[6], P. G. Conrad[1], C. Conley[7], S. Vitale[1], A. C. O'Brien[6], V. Riggi[1], K. Rogers[8]

Water-rock interactions are relevant to planetary habitability, influencing mineralogical diversity and the production of organic molecules. We examine carbonates and silicates in the martian meteorite Allan Hills 84001 (ALH 84001), using colocated nanoscale analyses, to characterize the nature of water-rock reactions on early Mars. We find complex refractory organic material associated with mineral assemblages that formed by mineral carbonation and serpentinization reactions. The organic molecules are colocated with nanophase magnetite; both formed in situ during water-rock interactions on Mars. Two potentially distinct mechanisms of abiotic organic synthesis operated on early Mars during the late Noachian period (3.9 to 4.1 billion years ago).

The martian meteorite Allan Hills 84001 (ALH 84001) formed during the Noachian period on Mars: It has an igneous crystallization age of ~4.09 billion years (*1–3*). ALH 84001 is predominately composed of the silicate mineral orthopyroxene (hereafter Opx). It also contains carbonate globules (*3*) that have been linked to aqueous processes on early Mars ~3.9 billion years ago (*1, 2*). As one of the oldest known rocks from Mars, ALH 84001 serves as a window into early planetary processes that may also have occurred on early Earth (*4*). Organic carbon, including possible nitrogen-containing organic compounds (*4–7*), has been described in ALH 84001. Hypotheses as to the provenance and formation mechanisms of these organics include abiotic production by impact-related (*8*), igneous (*6*), and/or hydrothermal processes (*5, 7*); biological production by putative ancient martian organisms (*4*); and terrestrial contamination (*9, 10*). To investigate the identity, origin, and formation mechanisms of organics, we applied colocated nanoscale spectral, imaging, structural, and isotopic analysis techniques to thin foils extracted from two subsamples of ALH 84001: the reported magnetite-rich crush zones of a thin section (designated ALH 84001,347) (*11*) and a cross-section through the center of a carbonate globule on a fresh fracture surface (designated ALH 84001,336) (*12*).

Colocated nanoscale analyses

We used a focused ion beam (FIB) to extract a foil (Fig. 1A) from an iron oxide–rich (magnetite) vein in a thin section of ALH 84001,347 (wider context shown in fig. S1C). Transmission electron microscopy (TEM) imaging of this foil shows that the Opx has a saw tooth or dentate appearance at its edge (Fig. 1, B and C); this is a characteristic feature of aqueous dissolution (*13*). The altered Opx surfaces are associated with a fibrous phase (labeled "1" in Fig. 1B) and an area of nanocrystalline material that is infilled with another fibrous phase (labeled "2" in Fig. 1C). Higher-resolution TEM images of the fibrous phases show that they are associated with nanocrystals of magnetite and carbonate (fig. S6B). Elemental compositional analyses of fibrous phases 1 and 2 show that their compositions are similar to those of lizardite and/or antigorite and to that of talc, respectively (Fig. 1D, table S1, and fig. S2). Bright-field TEM images of fibrous phases 1 and 2 (Fig. 1, E and G) indicate crystalline lattice fringes in an amorphous matrix with electron diffraction patterns (Fig. 1, F and H) exhibiting lattice *d*-spacings of ~9.6 Å and 9.2 to 9.7 Å (mean: 9.5 Å), respectively. These values correspond to basal plane distances of sheet silicates. We infer that both fibrous phases are predominately Fe-Mg silicates, containing a small amount of Al, with the appearance, chemical composition (Fig. 1D), and *d*-spacing characteristics of talc or possibly a serpentine subgroup mineral (Fig. 1D), hereafter referred to as a talc-like phase (*14–16*).

We obtained scanning transmission x-ray microscopy (STXM) spectral analysis of carbon (the C 1s edge) associated with the fibrous areas (Fig. 1I). These data indicate a range of organic functional groups colocalized with the fibrous phases (Fig. 1, B and C). The STXM peak distribution of these organics does not match potential contamination by the thin-section polymer used in ALH 84001,347 (table S2). The spectra indicate the presence of aromatic organic carbon (peak at 284.9 eV) and a range of organic oxygen functional groups, including carbonyl (286.5 eV), carboxyl (288.5 eV), and inorganic carbonate (290.4 eV). A small peak at 287.9 eV in spectrum 3 (asterisk in Fig. 1I) may be due to the presence of aliphatic or amidyl (C–NH) organic group functionality (spectrum 3 in Fig. 1I and table S2). We used nanoscale secondary ion mass spectrometry

[1]Carnegie Institution for Science, Earth and Planets Laboratory, Washington, DC 20015, USA. [2]Deutsches GeoForschungsZentrum, Telegrafenberg, 14473 Potsdam, Germany. [3]Department of Earth Sciences, Free University of Berlin, 12249 Berlin, Germany. [4]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK. [5]NASA Johnson Space Center, Houston, TX 77058, USA. [6]School of Geographical and Earth Science, University of Glasgow, Glasgow G12 8QQ, UK. [7]NASA Ames Research Center, Mountain View, CA 94035, USA. [8]Earth and Environmental Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA.
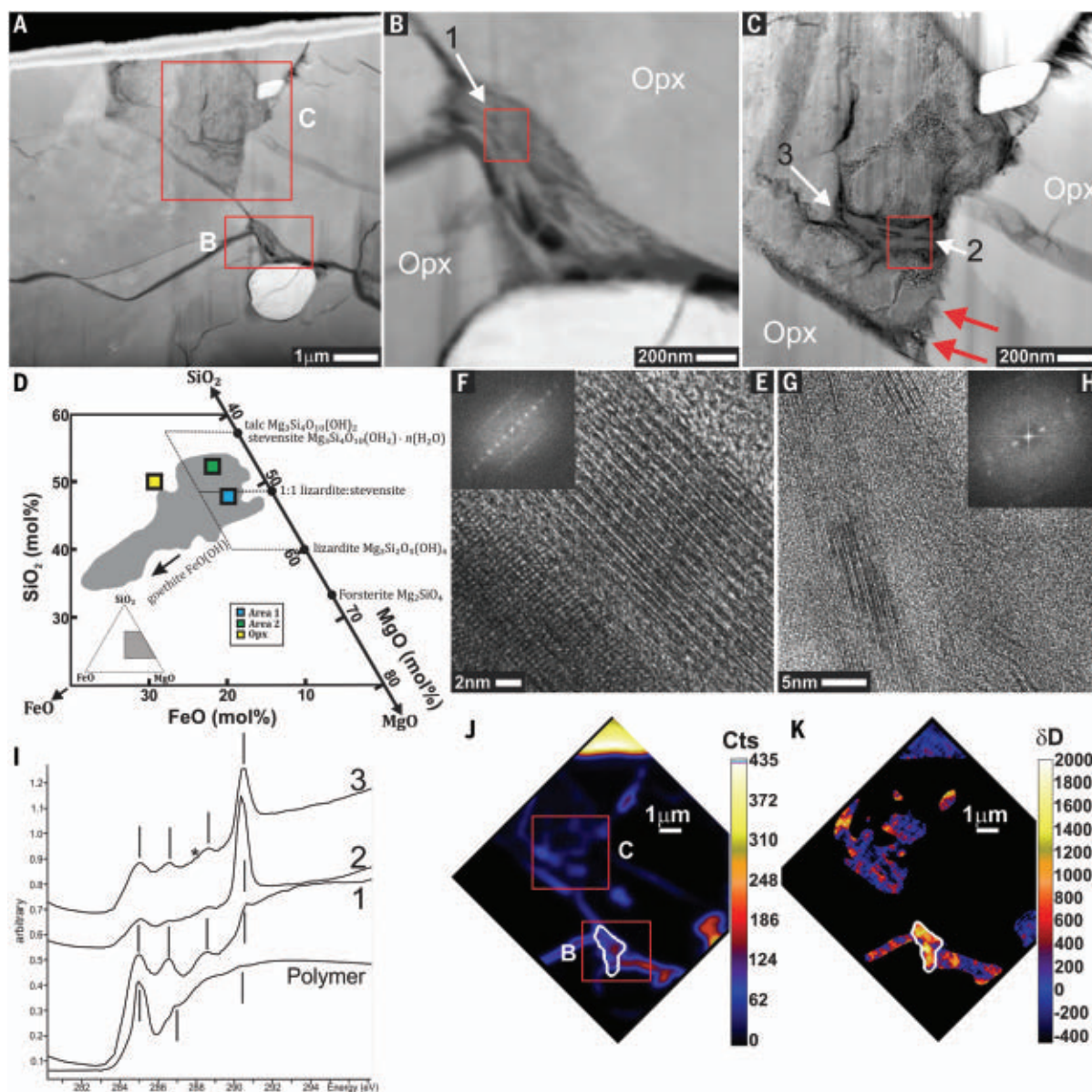*Corresponding author. Email: asteele@carnegiescience.edu

Fig. 1. Analysis of FIB foil taken from the ALH 84001 matrix. (**A**) Overview image of the foil; wider context is shown in fig. S1C. Red boxes outline the areas depicted in (B) and (C). (**B**) High-angle annular dark-field (HAADF) TEM image showing a fibrous phase infilling a crack between two slightly weathered Opx grains (arrow labeled "1"). The red box indicates area of analyses depicted in (E). (**C**) TEM HAADF image of microcrystalline material (white arrow labeled "3") infilled with a fibrous material (white arrow labeled "2"). The red arrows indicate areas of denticle formation from pyroxene dissolution, and the red box marks the area of analyses depicted in (G). (**D**) Truncated portion of a MgO-SiO$_2$-FeO ternary diagram, which shows the MgO, FeO, and SiO$_2$ molar abundances measured by energy-dispersive x-ray spectroscopy in areas 1 and 2 and host Opx [from white arrows in (B) and (C)] compared to those of serpentines from Oman (gray shaded region) (*14*). The inset shows the location of the truncated area in gray (see fig. S2B for a full ternary diagram). (**E**) TEM image of crystallites from the area within the red box in (B), showing lattice fringes indicative of a sheet silicate. (**F**) Fast Fourier transform (FFT) diffraction pattern of (E). (**G** and **H**) Same as (E) and (F) but for the area within the red box in (C). (**I**) STXM spectra from areas labeled with white arrows 1 to 3 in (B) and (C) compared with the spectrum for a thin-section polymer within the sample. Black vertical lines represent peak centers of interest; the asterisk denotes a small shoulder peak (see main text for details). (**J**) NanoSIMS map of ^{12}C abundance (Cts, counts); the red boxes indicate the fibrous material depicted in (B) and (C). (**K**) Same as (J) but for H isotopic analysis. The area outlined in white corresponds to the fibrous material in (B) and has δD = 790 ± 140‰.

(NanoSIMS) to confirm the presence of carbon-rich material in the fibrous areas (Fig. 1J). Hydrogen isotopic analyses of these carbon-rich areas indicate that the organic matter and fibrous material are enriched in deuterium relative to Earth, with deuterium excess (δD) values of ~790 ± 140 per mil (‰) (Fig. 1K). This range of

δD values differs from that of the thin-section polymer, which we measured to be ~−250 ± 50‰. The ALH 84001 data have a δD value within the range of the martian crustal H$_2$O reservoir, which is ~700 to 2700‰ (*17*). The high, positive δD values, combined with the multitude of organic carbon functional groups,

indicate negligible contamination of the talc-like phase by the thin-section polymer.

We cut additional FIB sections across a carbonate globule from a fresh fracture surface of ALH 84001,336 to evaluate a portion of the meteorite that had not come into contact with the thin-section polymer (fig. S3). A FIB slice
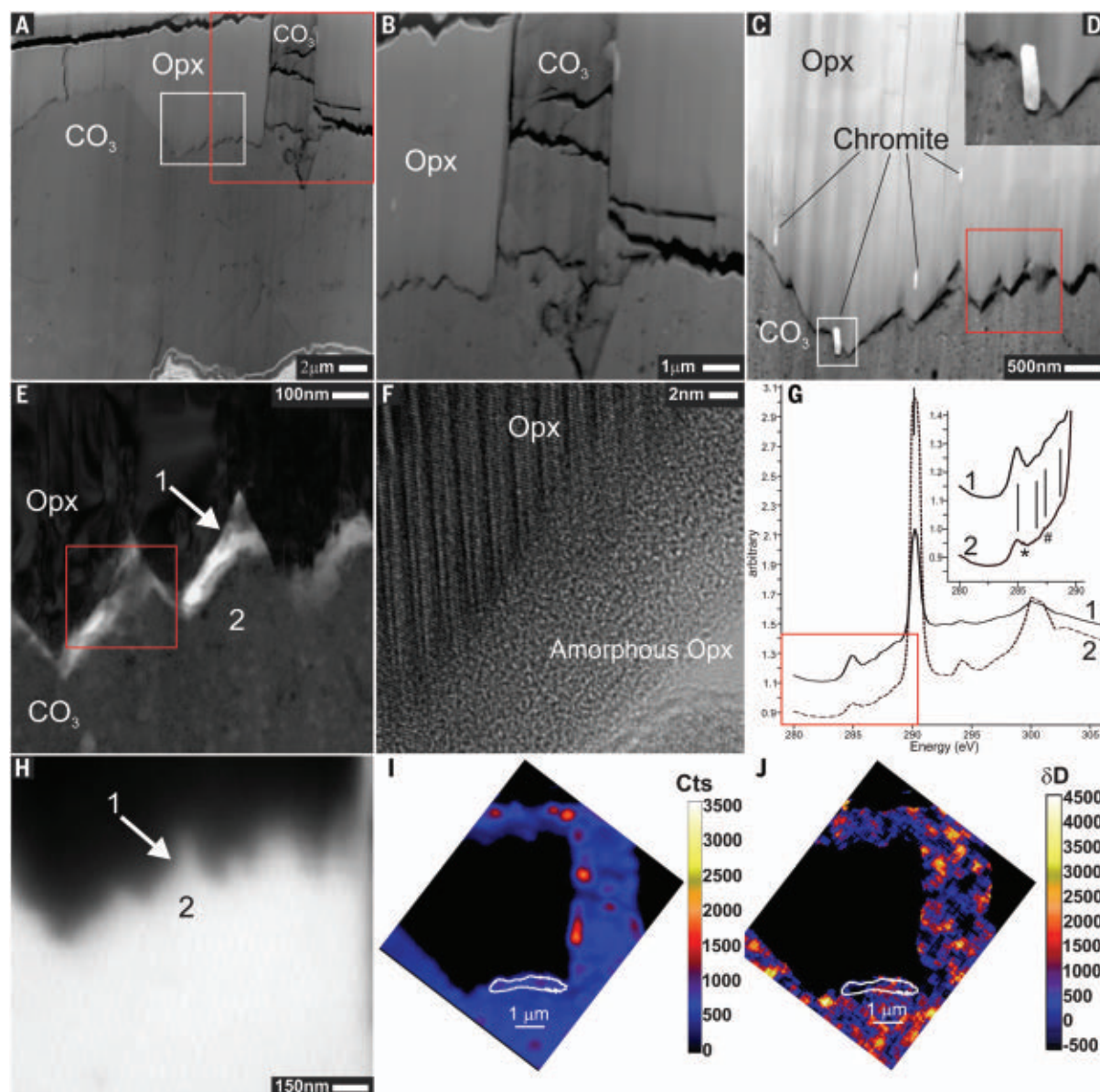
Fig. 2. Analysis of FIB foil taken from the center of a carbonate globule from ALH 84001. (**A**) TEM HAADF image of a slice through the center of the globule. The red box indicates the area shown in (B); the white box indicates the area depicted in (C) and (D). Broader context is shown in fig. S3. (**B**) Zoomed-in HAADF image [from red box in (A)] showing a carbonate vein connected to the carbonate globule. (**C**) HAADF image zoomed in on the carbonate-pyroxene contact; multiple small chromite grains oriented along the 110 axis of Opx are visible. The Opx exhibits denticular texture indicative of aqueous weathering. The white box surrounds a small chromite grain [magnified in (D)], and the red box shows a denticular area of Opx [magnified in (E)]. (**D**) Close-up image of a chromite grain [from (C)] protruding into the carbonate from the pyroxene. (**E**) TEM bright-field image of the carbonate-pyroxene contact shown in (C). The red box indicates the approximate area shown in (F). (**F**) TEM lattice fringe image of the crystalline-amorphous Opx interface, as indicated by the presence of lattice fringes and an amorphous material between the carbonate and the pyroxene. (**G**) STXM spectra of regions 1 and 2 from (E) and (H). The black lines indicate major peaks, the asterisk denotes a small shoulder peak, and the # symbol indicates a possible artefact peak. The inset is a magnified view of the region within the red box. (**H**) STXM map of the carbonate peak at 290 eV of the same denticular area as that shown in the red box in (C). The locations of STXM spectra 1 and 2 are indicated. (**I**) NanoSIMS ^{12}C map of the carbonate vein shown in (B) and (C). The area outlined in white corresponds to the Opx-carbonate contact in the same region as that depicted in (C) and (E). (**J**) NanoSIMS δD map of the same area as depicted in (I). For the organic-rich areas measured within the region outlined in white, we measured a δD of 850 ± 400‰.

from the center of the carbonate globule reveals a denticular interface between the carbonate globule and Opx (Fig. 2, A to E, and fig. S2C), similar to that shown in Fig. 1. This Opx dissolution texture appears to have propagated along the (110) cleavage plane of Opx,

perpendicular to (what appear to be) shock lamellae along the (100) plane (fig. S5). Chromite grains are present throughout the Opx (Fig. 2D). One such grain appears to have resisted the alteration process of the Opx as it protrudes from Opx into the carbonate

(Fig. 2D). In addition to alteration at the Opx-globule interface, the FIB slice revealed a carbonate vein within the Opx that is contiguous with the globule (Fig. 2B). The carbonate within the vein and the globule appears to be zoned in a similar way to that described previously
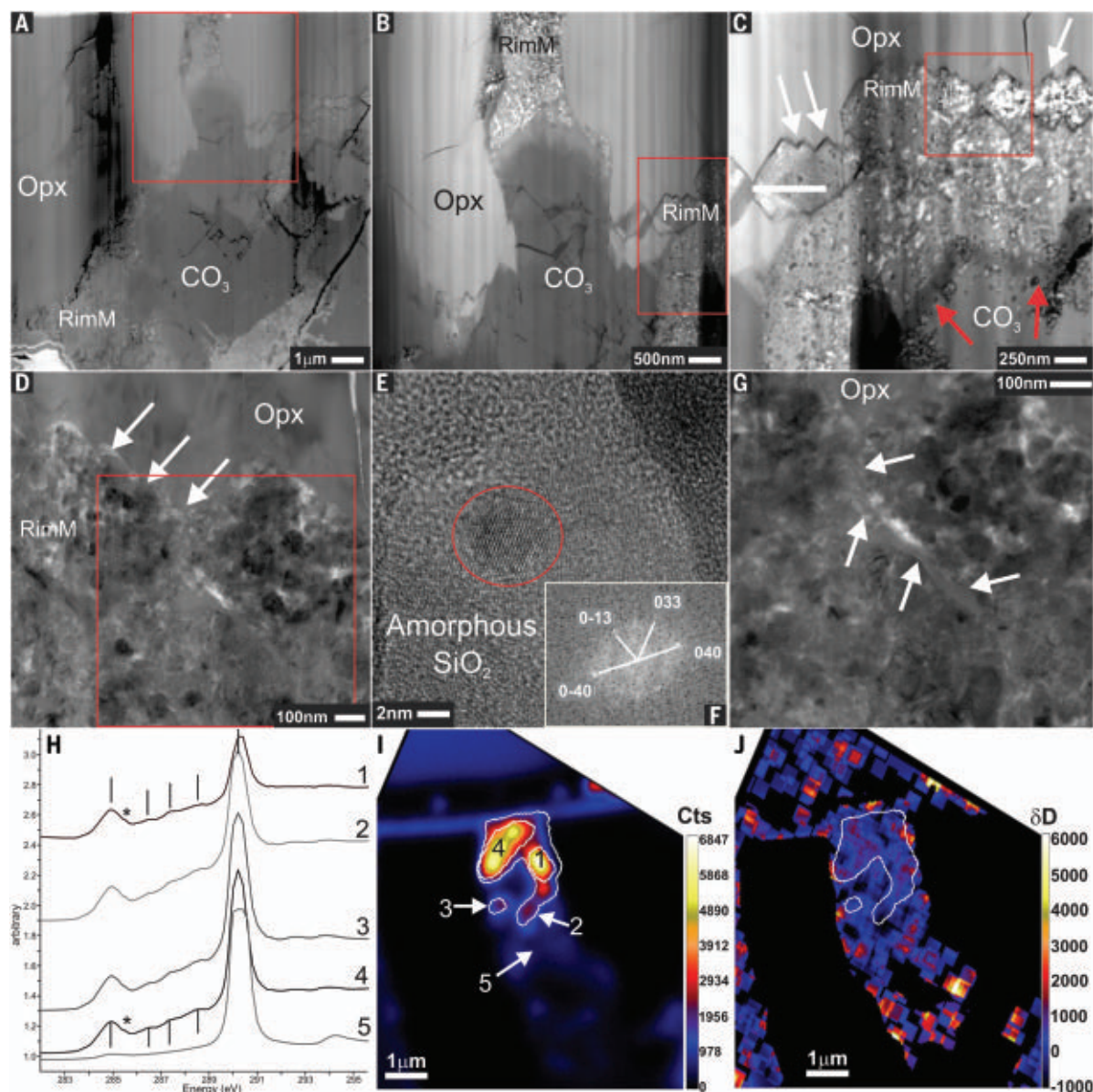
Fig. 3. Analysis of FIB foil taken from the rim of a carbonate globule from ALH 84001. The carbonate globule is the same as that shown in Fig. 2. (**A**) TEM HAADF image of a FIB section taken from the rim area of the globule shown in fig. S3 (the top of the globule is at the bottom of the image). The red box indicates the area shown in (B) and approximates the areas in (I) and (J). (**B**) TEM HAADF image of a vein of CO_3 and RimM material at the base of the carbonate globule. Rim material can also be observed underneath the carbonate at the pyroxene-carbonate interface, indicated by the red box, which is shown in more detail in (C). (**C**) Close-up view of the rim material at the interface between Opx and the CO_3 globule. White arrows denote the denticular boundary of the pyroxene edge; red arrows indicate a chevron-like contact between the carbonate and the RimM that we interpret as an area of transition or alteration between the pyroxene and carbonate. The red box denotes the area shown in (D). (**D**) Bright-field image of the area outlined by the red box in (C), revealing the contact of RimM and the CO_3 globule. The lighter areas in this image are holes or pores in the sample. White arrows indicate the boundary between Opx and RimM material. (**E**) TEM lattice fringe image of amorphous silica vein from (D) and (G), revealing a single magnetite grain (outlined by red ellipse) within a matrix of amorphous silica (marked as amorphous SiO_2). (**F**) FFT diffraction pattern of the magnetite grain in (E) [white arrows and text represent Miller indices (*hkl*)]. (**G**) Bright-field TEM image of the area outlined by the red box in (D). White arrows point to an amorphous silica vein that runs through RimM. (**H**) STXM spectra from spots taken within RimM and carbonate, showing differences in the amount of included organic matter within carbonate. Numbers indicate spectra taken from the areas shown in (I). The black vertical lines indicate major peaks; the asterisk denotes a small shoulder peak. (**I**) ^{12}C NanoSIMS map of the carbonate vein in (A) [approximate area shown is enclosed within the red box in (A)], depicting bright spots of organic material within carbonate. Numbers correspond to the spectra in (H). (**J**) δD map of the same area as in (I). The organic material within the white contour has δD = 400 ± 170‰.

(*18*), with voids throughout its structure that contain single magnetite or siderite nanocrystals (fig. S4). A bright-field TEM image of the Opx-globule interface is nanocrystalline in appearance (Fig. 2E) and filled with an amorphous silicate, which we identified by means of energy-dispersive x-ray spectroscopy analyses and by a lack of lattice fringes in high-resolution TEM imaging (Fig. 2F and fig. S6). STXM analysis of this material (Fig. 2G) shows peaks corresponding to aromatic/olefinic carbon, as well as carbonyl, carboxyl, and carbonate groups

(table S2). The carbonate peak at 290.4 eV dominates these spectra, but lower-energy organic carbon peaks are also present (Fig. 2G). Figure 2H shows a STXM map of the carbonate peak (at 290 eV) in the area of Opx dissolution textures shown in Fig. 2D, and the locations from which STXM spectra were collected are shown in Fig. 2, E and H. Hydrogen isotopic analysis of the organic carbon, measured in the same area as the STXM analysis (Fig. 2J), indicates that $\delta D = 850 \pm 400$‰, within the range of values for the martian crust (17).

Analysis of a third FIB section (Fig. 3), taken from the rim of the same carbonate globule (Fig. 2), revealed a chaotic interface between carbonate, Opx, and a magnetite- and pyrite-rich rim material (RimM) that occurs at the edge and at the base of the globule (Fig. 3B). A second veinlike feature that cuts across the Opx contains rim material that transitions into a carbonate vein that is contiguous with the globule (Fig. 3, B and C). A bright-field TEM image of the rim material (Fig. 3D) shows a nanocrystalline mixture of phases infilling the space between the Opx denticles. A single nanocrystal of magnetite (~5 nm in diameter; Fig. 3, E and F) is surrounded by an amorphous material composed mainly of silica (fig. S7). The C 1s STXM spectra of five different locations within the RimM material (Fig. 3H) reveal the presence of carbonate and organic carbon that exhibit a functionality distribution similar to that observed in the areas depicted in Figs. 1 and 2 (see also Fig. 3H and table S2). NanoSIMS analysis of ^{12}C and δD in this organic material (Fig. 3, I and J) also shows elevated δD values (400 ± 170‰) that are between those of the martian crust and depleted martian mantle (17).

Implications for early Mars

The microdenticular texture of Opx that we observe in ALH 84001 indicates aqueous and/or hydrothermal alteration of Opx, consistent with previous interpretations (13, 18). The carbonate, amorphous silicate, and hydrated silicate phases that formed during alteration of ALH 84001 are not volumetrically dominant, indicating that fluid-rock reactions were not active for a prolonged time period. However, the phase assemblages that occur at the outer edges of the altered Opx (i.e., amorphous silica; talc-like phases; magnetite; and predominately Fe-, Mg-, and Ca-rich carbonates) are similar to those observed in rocks from Earth that have undergone serpentinization and/or mineral carbonation, which indicates that similar alteration processes operated on early Mars. Whether the hydration and carbonation processes occurred as part of a single metasomatic event or as distinct events remains unclear. Previous work has suggested that ALH 84001 may have undergone more than one fluid event, possibly associated with

pre-ejection impact events that caused the formation of crush zones in ALH 84001 (11, 19, 20). Formation of serpentinites on Earth requires a balance of many variables (14), including host rock composition (21), temperature, pressure, pH, silica activity, anion and cation concentrations, reduction and oxidation, water activity in the fluid phase, and partial pressure of CO_2 (22–25). This balance can produce a diverse mix of talc- or serpentine-like mineral phases during aqueous and/or hydrothermal alteration (14, 21–25), making it difficult to constrain the conditions of fluid composition or the formation temperature for the alteration phases in ALH 84001, particularly given the later influence of impact ejection (12). Consequently, we cannot determine the timing or formation mechanisms of the talc, carbonate, and organics, but the simplest explanation is that these materials are reaction products of the interaction of Opx with hydrothermal fluid(s) of neutral to high pH (22). Our observations of Opx corrosion, the colocation of organic carbon and nanophase magnetite, and the alteration mineral assemblage(s) lead us to conclude that serpentinization and carbonation reactions occurred on Mars in the late Noachian and were a source of endogenous martian organic carbon.

Organic synthesis

Serpentinization is an abiotic organic synthesis mechanism whereby basaltic rocks react with an aqueous fluid, producing serpentine minerals, magnetite, and hydrogen (26, 27). The hydrogen produced in this reaction is then available to reduce aqueous CO_2 to methane (via the Sabatier reaction or the reverse water-shift reaction) as well as to CO and other organics such as formic acid and formaldehyde (26–29). CO and H can also react via Fischer-Tropsch–type reactions to produce alkanes and other organic molecules, including nitrogen-containing organics (29, 30). In ALH 84001, organics are colocated with magnetite in two different mineral assemblages. In the portion of the sample shown in Fig. 1, the magnetite coexists with a talc-like phase, indicating that martian serpentinization reactions are responsible for the formation of the observed organic compounds. By contrast, portions of the sample depicted in Figs. 2 and 3 show the presence of magnetite within an area containing only amorphous silica, carbonate, and organic carbon, indicating that martian mineral carbonation reactions are also responsible for the formation of organics. We did not detect organic material in cracks or fissures outside of those mineral assemblages; therefore, we discount external sources of organic material formed or transported into the sample from elsewhere on Mars. Previous studies have suggested biogenic processes (4), thermal decomposition of siderite (8), and/or aqueous abiotic hydrothermal processes (5, 7) for the origin of

organics in ALH 84001. Our results indicate that the deuterium-rich organic material is associated with the precipitation of nanophase magnetites in the presence of silica and carbonate (akin to mineral carbonation) and in the presence of talc-like phases (akin to serpentinization), consistent with aqueous abiotic hydrothermal processes (26–29). The organic material is aromatic and associated with carbonyl, carboxyl, and carbonate group functionality; similar features have been observed to be spatially correlated with mineral catalysts in other martian meteorites (e.g., Tissint, Nakhla, and NWA 1950) (31). Similar refractory aromatic organic material has also been detected through in situ analyses facilitated by the Curiosity rover (32).

Martian habitability

We conclude that aqueous alteration of the Opx in ALH 84001 caused the formation of carbonate globules, amorphous silicates and silica, and talc-like phases. This phase assemblage indicates that serpentinization and mineral carbonation reactions took place on early Mars, consistent with orbital observations of serpentinized terrains on Mars and studies of the martian meteorite Nakhla (33–35). Serpentinization and mineral carbonation reactions occur by aqueous alteration of mafic igneous mineral assemblages by CO_2-bearing fluids. Organic synthesis in ALH 84001 proceeded in a manner similar to serpentinization of rocks from Earth, producing aromatic, aliphatic, carbonyl, carboxyl, and carbonate species from aqueous CO_2. The correlation of organic material with both serpentinization and carbonation assemblages indicates that abiotic organic synthesis can occur from both reactions. The similarity of the organics in the ~4.0-billion-year-old ALH 84001 and those found in the ~600-million-year-old Tissint meteorite (31) indicate that Mars hosted abiotic organic synthesis reactions for much of its history. On Earth, these reactions are responsible for abiotic organic synthesis, methane production, and mineralogical diversity (29). On Mars, such reactions are relevant to habitability and have been invoked to explain the presence of methane in the atmosphere (36).

REFERENCES AND NOTES

1. L. E. Borg et al., Science 286, 90–94 (1999).
2. T. J. Lapen et al., Science 328, 347–351 (2010).
3. D. W. Mittlefehldt, Meteoritics 29, 214–221 (1994).
4. D. S. McKay et al., Science 273, 924–930 (1996).
5. A. Steele et al., Meteorit. Planet. Sci. 42, 1549–1566 (2007).
6. A. Steele et al., Science 337, 212–215 (2012).
7. M. Koike et al., Nat. Commun. 11, 1988 (2020).
8. A. H. Treiman, Astrobiology 3, 369–392 (2003).
9. T. Stephan, E. K. Jessberger, C. H. Heiss, D. Rost, Meteorit. Planet. Sci. 38, 109–116 (2003).
10. A. Steele et al., Meteorit. Planet. Sci. 35, 237–241 (2000).
11. A. H. Treiman, Meteoritics 30, 294–302 (1995).
12. Materials and methods and supplementary text are available as supplementary materials.
13. C. M. Phillips-Lander, C. Legett, A. S. Elwood Madden, M. E. Elwood Madden, Am. Mineral. 102, 1915–1921 (2017).

14. J. C. de Obeso, P. B. Kelemen, *Philos. Trans. R. Soc. A* **378**, 20180433 (2020).
15. D. B. Nahon, F. Colin, *Am. J. Sci.* **282**, 1232–1243 (1982).
16. E. J. W. Whittaker, J. Zussman, *Mineral. Mag. J. Mineral. Soc.* **31**, 107–126 (1956).
17. J. J. Barnes *et al.*, *Nat. Geosci.* **13**, 260–264 (2020).
18. K. L. Thomas-Keprta, S. J. Clemett, D. S. McKay, E. K. Gibson, S. J. Wentworth, *Geochim. Cosmochim. Acta* **73**, 6631–6677 (2009).
19. C. M. Corrigan, R. P. Harvey, *Meteorit. Planet. Sci.* **39**, 17–30 (2004).
20. J. M. Eiler, J. W. Valley, C. M. Graham, J. Fournelle, *Geochim. Cosmochim. Acta* **66**, 1285–1303 (2002).
21. B. W. Evans, *J. Petrol.* **49**, 1873–1887 (2008).
22. A. N. Paukert, J. M. Matter, P. B. Kelemen, E. L. Shock, J. R. Havig, *Chem. Geol.* **330–331**, 86–100 (2012).
23. L. E. Mayhew, E. T. Ellison, T. M. McCollom, T. P. Trainor, A. S. Templeton, *Nat. Geosci.* **6**, 478–484 (2013).
24. D. Daval *et al.*, *Geochim. Cosmochim. Acta* **74**, 2615–2633 (2010).
25. C. Boschi, A. Dini, L. Dallai, G. Ruggieri, G. Gianelli, *Chem. Geol.* **265**, 209–226 (2009).
26. N. G. Grozeva, F. Klein, J. S. Seewald, S. P. Sylva, *Philos. Trans. R. Soc. A* **378**, 20180431 (2020).
27. T. M. McCollom, J. S. Seewald, *Chem. Rev.* **107**, 382–401 (2007).
28. B. Sherwood Lollar, T. D. Westgate, J. A. Ward, G. F. Slater, G. Lacrampe-Couloume, *Nature* **416**, 522–524 (2002).
29. N. G. Holm, A. Neubeck, *Geochem. Trans.* **10**, 9 (2009).
30. M. Schulte, E. Shock, *Meteorit. Planet. Sci.* **39**, 1577–1590 (2004).
31. A. Steele *et al.*, *Sci. Adv.* **4**, eaat5118 (2018).
32. J. L. Eigenbrode *et al.*, *Science* **360**, 1096–1101 (2018).
33. C. E. Viviano, J. E. Moersch, H. Y. McSween, *J. Geophys. Res. Planets* **118**, 1858–1872 (2013).
34. B. L. Ehlmann, J. F. Mustard, S. L. Murchie, *Geophys. Res. Lett.* **37**, L06201 (2010).
35. T. Tomkinson, M. R. Lee, D. F. Mark, C. L. Smith, *Nat. Commun.* **4**, 2662 (2013).
36. J. R. Lyons, C. Manning, F. Nimmo, *Geophys. Res. Lett.* **32**, L13201 (2005).
37. A. Steele *et al.*, Data for "Organic synthesis associated with serpentinization and carbonation on early Mars," Dryad (2021); https://doi.org/10.5061/dryad.c2fqz6198.

PLANT SCIENCE

An RNA exosome subunit mediates cell-to-cell trafficking of a homeobox mRNA via plasmodesmata

Munenori Kitagawa[1], Peipei Wu[1], Rachappa Balkunde[1]†, Patrick Cunniff[1], David Jackson[1,2]*

Messenger RNAs (mRNAs) function as mobile signals for cell-to-cell communication in multicellular organisms. The KNOTTED1 (KN1) homeodomain family transcription factors act non–cell autonomously to control stem cell maintenance in plants through cell-to-cell movement of their proteins and mRNAs through plasmodesmata; however, the mechanism of mRNA movement is largely unknown. We show that cell-to-cell movement of a KN1 mRNA requires ribosomal RNA–processing protein 44A (AtRRP44A), a subunit of the RNA exosome that processes or degrades diverse RNAs in eukaryotes. AtRRP44A can interact with plasmodesmata and mediates the cell-to-cell trafficking of KN1 mRNA, and genetic analysis indicates that AtRRP44A is required for the developmental functions of SHOOT MERISTEMLESS, an *Arabidopsis* KN1 homolog. Our findings suggest that AtRRP44A promotes mRNA trafficking through plasmodesmata to control stem cell–dependent processes in plants.

C ell-to-cell communication in multicellular organisms promotes cell fate specification and coordination of development. As one way to transmit information between cells, plants selectively traffic transcription factors through plasmodesmata, cell wall–embedded channels that connect the cytoplasm of neighboring cells (*1–3*). The maize KNOTTED1 (KN1) homeodomain transcription factor was the first mobile protein found to use this trafficking pathway (*4*). Previous studies have identified regulators of KN1 protein trafficking (*5–7*); however, KN1 protein traffics with its mRNA (*4*). Selective trafficking of mRNAs in plants is prevalent (*8–12*); however, the mechanism by which this occurs through plasmodesmata has not been addressed. Here, we identify *Arabidopsis* ribosomal RNA–processing protein 44A (AtRRP44A) as an essential factor for the cell-to-cell trafficking of KN1 mRNA and show that AtRRP44A-dependent mRNA trafficking is required for cell-to-cell protein trafficking and stem cell functions in plants.

Isolation of KN1 trafficking mutants

We previously established a genetic screen in *Arabidopsis* to identify regulators of KN1 cell-to-cell trafficking by using a "trichome rescue system" (*7, 13*). Trichomes are hairlike extensions of the leaf epidermis. Their development requires the cell-autonomous activity of GLABROUS1 (GL1), a MYB transcription factor (*14*). In our system, a fusion protein of green fluorescent protein (GFP), GL1, and the KN1 C-terminal trafficking domain (KN1^C) is expressed in the mesophyll cell layers of

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. ²National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, P.R. China.
*Corresponding author. Email: jacksond@cshl.edu
†Present address: Bayer Crop Science LLC, Chesterfield, MO 63017, USA.

leaves of trichomeless *gl1* mutants by using the Rubisco small subunit 2b promoter (*pRbcS::GFP~GL1~KN1^C*) (*7*). Trafficking of GFP~GL1~KN1^C to the epidermis rescues trichome formation in this line; thus, trichome number is an output for KN1 trafficking. In an ethyl methanesulfonate (EMS) mutagenesis screen of trichome rescue lines, we isolated two mutants that lacked trichomes, which were initially referred to as *rb31-7* and *mk5-140* (Fig. 1, A to C). Consistent with the loss of trichomes, epidermal GFP~GL1~KN1^C accumulation decreased significantly in the mutants (Fig. 1, D to F) despite similar expression compared with that of the parental control trichome rescue lines in mesophyll cell layers (Fig. 1, G to I). These observations suggest that both mutants reduced the trafficking of GFP~GL1~KN1^C from the mesophyll cell layers to the epidermis. To investigate whether trichome loss in the mutants was simply due to a reduction in transgene expression, we measured GFP~GL1~KN1^C fluorescence in mesophyll cells of the mutants. We found that it was ~50 to 70% of the level in the parental trichome rescue lines (fig. S1A). However, this reduction was not the cause of trichome loss because plants hemizygous for the trichome rescue transgene also had ~50% expression, and this was sufficient for trichome rescue (fig. S1, A and B). Thus, we confirmed that KN1 trafficking was inhibited in *rb31-7* and *mk5-140* mutants.

KN1 trafficking mutants encode AtRRP44A

Both *rb31-7* and *mk5-140* mutants behaved as single recessive loci. We mapped them by sequencing M3 pools of mutants or nonmutant siblings to ~44× coverage. Using the MutMap+ pipeline (*15*), we identified potentially causal point mutations of *rb31-7* and *mk5-140* within the same gene, At2g17510 [guanine (G) to adenine (A) and cytosine (C) to thymine (T), which causes disruptive Cys^{551} to Tyr and Pro^{781}

to Leu amino acid substitutions, respectively] (Fig. 1, J and K). At2g17510 encodes AtRRP44A, which is a subunit of the RNA exosome. We confirmed this as the causal gene by complementing trichome rescue defects in both mutants by using a TagRFP-T (hereafter RFP, red fluorescent protein) fusion protein that was expressed by its native regulatory sequences (*pAtRRP44A::AtRRP44A~RFP*) (Fig. 1, L and M), so we renamed our alleles as *atrrp44a-4* and *atrrp44a-5*, respectively. Null alleles of *AtRRP44A* in *Arabidopsis* are lethal (*16*, *17*), as in yeast, *Drosophila*, and human cells (*18–21*); thus, ours are likely to be weak alleles. The RNA exosome is a multiprotein complex that is involved in the processing and degradation of a wide range of RNAs in eukaryotes, and RRP44 is a catalytic subunit with 3′-to-5′ exoribonuclease and endoribonuclease activities (*22–24*). Two conserved domains, RNB (ribonuclease B) and PIN (pilT N-terminus), are responsible for these activities, respectively (Fig. 1K). The mutations in *atrrp44a-4* and *atrrp44a-5* were in highly conserved residues in the RNB domain (Fig. 1K and fig. S2A) (*25–27*) in the vicinity of the RRP44 catalytic center (fig. S2B) (*28*), which supports the notion that they are important for function.

atrrp44a mutants enhance *stm* and *cct8* phenotypes

KN1 can fully complement the shoot meristem defects of its *Arabidopsis* homolog SHOOT MERISTEMLESS (STM), whose mobility is essential for meristem maintenance (*13*, *29*, *30*). *atrrp44a-5* mutants also blocked trichome rescue by a *GFP~GL1~STM*[C] fusion, suggesting that AtRRP44A also participates in STM trafficking (fig. S3). Thus, we next investigated whether AtRRP44A affects STM function. *atrrp44a* mutants had no obvious defects in shoot meristem size (Fig. 2, A, B, and E, and fig. S4A); however, they significantly enhanced the meristem phenotypes of *stm-10*, a weak allele of *stm* (*31*). Shoot meristems of *atrrp44a-4; stm-10* double mutants were significantly smaller than *stm-10* single mutants (Fig. 2, A to E) and terminated in many of the *atrrp44a-5; stm-10* double-mutant seedlings (22 out of 24) (fig. S4B). Similarly, meristems arrested in many of the *atrrp44a-4; stm-10* double mutants (fig. S4C), which resulted in seedlings with fewer leaves (fig. S4D). Consistent with these phenotypes, the expression of *STM* and of additional shoot meristem genes *CLAVATA3* (*CLV3*) and *WUSCHEL* appeared disorganized and/or reduced in the *atrrp44a; stm-10* double-mutant meristems, and *CLV3* expression was significantly lower in quantitative reverse transcription polymerase chain reaction (RT-qPCR) experiments (fig. S4, E and F). Collectively, our data suggest that AtRRP44A is required for STM-dependent shoot meristem maintenance,
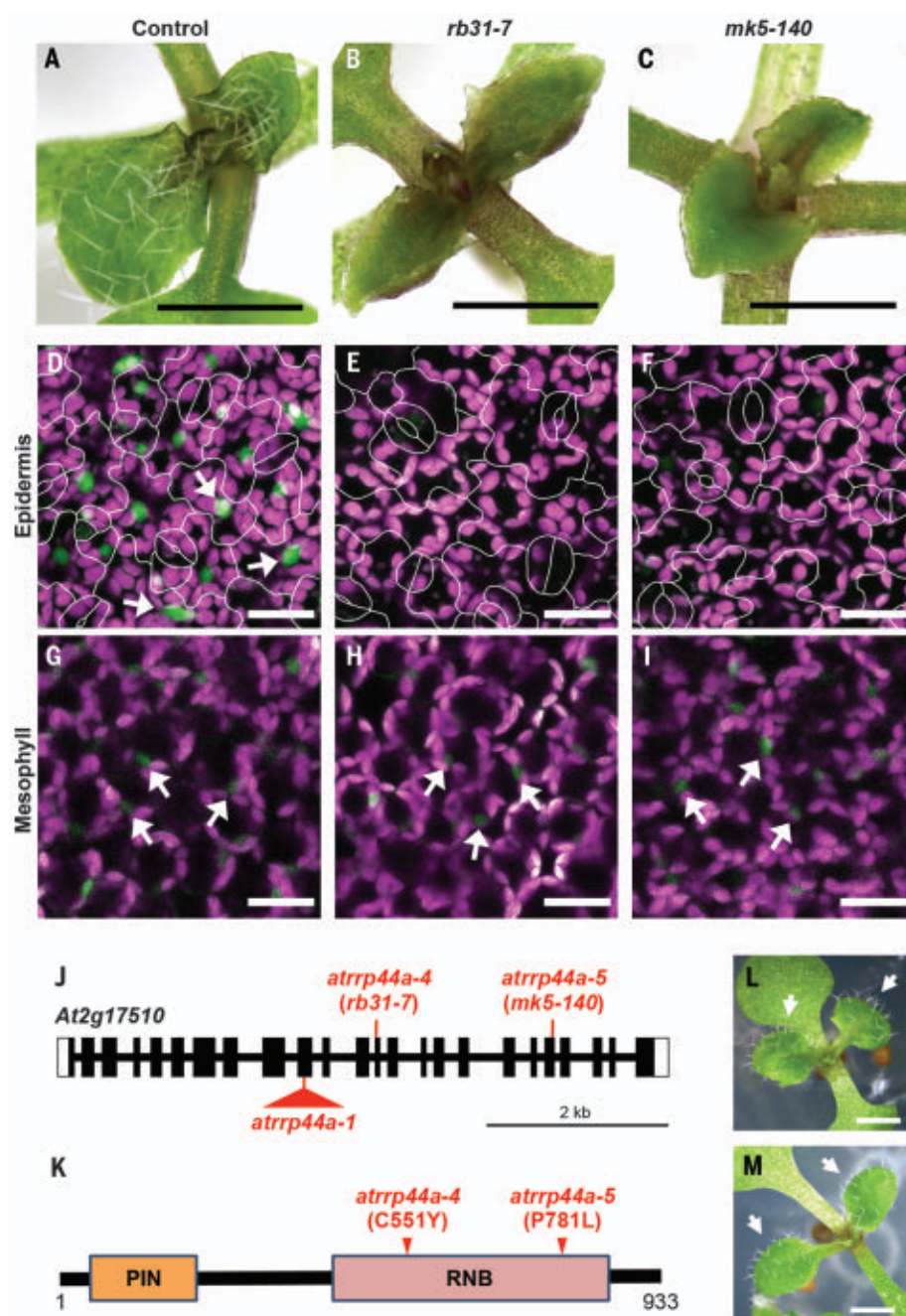
Fig. 1. AtRRP44A promotes KN1 trafficking. (**A** to **C**) Starting line for EMS mutagenesis (A) expresses GFP~GL1~KN1[C] in mesophyll cells in a *gl1* mutant background and rescues trichome formation, whereas the *rb31-7* (B) and *mk5-140* (C) mutants lost trichome rescue. Scale bars, 1 mm. (**D** to **I**) Confocal images of a true leaf from the starting trichome rescue line show the accumulation of GFP~GL1~KN1[C] in both the epidermis (D) and the mesophyll (G), whereas the *rb31-7* and *mk5-140* mutants have a severe reduction of GFP~GL1~KN1[C] in the epidermis [(E) and (F), respectively] but normal accumulation in the mesophyll [(H) and (I), respectively]. The arrows show representative nuclei with GFP~GL1~KN1[C] fusion protein. White lines show the outlines of epidermal cells. Chlorophyll autofluorescence is shown in magenta. Scale bars, 20 μm. (**J**) *AtRRP44A* (At2g17510) gene structure, showing different mutant alleles, *atrrp44a-1*, *atrrp44a-4* (*rb31-7*), and *atrrp44a-5* (*mk5-140*). Black bars indicate exons; lines and white boxes represent introns and untranslated regions, respectively. (**K**) AtRRP44A protein structure, showing different domains and the mutation sites in *atrrp44a-4* and *atrrp44a-5* in the RNB domain. (**L** and **M**) *atrrp44a-4* and *atrrp44a-5* mutants rescued by native expression of RFP-tagged AtRRP44A have normal trichome rescue (arrows). Scale bars, 1 mm.

perhaps by controlling STM RNA trafficking. To further support a role for *atrrp44a* in meristem development, we attempted to enhance our weak alleles by combining in a hemizygous state with the *atrrp44a-1* null allele (Fig. 1J) (*17*). Indeed, these combinations significantly reduced the shoot meristem size (fig. S4G). We also assayed genetic interactions between *AtRRP44A* and *CCT8*, a regulator of KN1 trafficking (*7*) that is required for STM function in stem cell maintenance. The *atrrp44a; cct8-1* double mutants developed dwarf shoots with fasciated stems [*n* = 20 of 20 plants (*atrrp44a-4; cct8-1*); 18 of 18 plants (*atrrp44a-5; cct8-1*)] (Fig. 2, F and G, and fig. S5). The double mutants also occasionally made nonfasciated shoot branches, which terminated prematurely [*n* = 10 of 20 plants (*atrrp44a-4; cct8-1*); 11 of 18 plants (*atrrp44a-5; cct8-1*)] (Fig. 2H and fig. S5B). These phenotypes were not observed in the single mutants, suggesting that AtRRP44A regulates the balance of proliferation and differentiation in the shoot meristem through a chaperonin-dependent pathway. Collectively, our data support the idea that AtRRP44A is required for the function and trafficking of KN1 and STM, which are essential for meristem maintenance.

AtRRP44A can localize to plasmodesmata

Given our hypothesis that AtRRP44A promotes cell-to-cell trafficking of STM, we next examined whether its expression overlaps with STM in the shoot meristem. We transformed the native AtRRP44A~RFP fusion construct into plants that were heterozygous for the *atrrp44a-1* null allele. This fusion construct was functional, as it complemented the lethality of the null allele in the subsequent generation (fig. S6A). AtRRP44A~RFP expression was observed throughout the shoot meristem and flower primordia (Fig. 3A and fig. S6B), which overlapped with STM expression (*32*), as well as in epidermal and mesophyll layers of leaf primordia (Fig. 3B), where GFP~GL1~KN1^{C} traffics in the trichome rescue system. Thus, AtRRP44A was expressed broadly, including in the shoot meristem where STM traffics. The fusion protein accumulated in nuclei, as expected, mainly in the nucleolus in meristem cells and throughout the nucleoplasm in leaf primordia (Fig. 3, A and B) (*16*).

We next investigated how AtRRP44A might facilitate trafficking. Previous studies that used microinjection or the trichome rescue system indicated that KN1 protein promotes the trafficking of its mRNA (*4, 13*), suggesting that KN1 protein and mRNA traffic as an mRNA–protein (mRNP) complex. AtRRP44A binds to the exosome core complex, which is involved in the processing of rRNAs, mRNAs, and non-coding RNAs in *Arabidopsis* (*17, 33, 34*), but whether AtRRP44A participates in mRNA

degradation or processing is unclear (*17*). Nonetheless, because AtRRP44A is a ribonuclease, we assessed whether it might affect trafficking indirectly by degradation or processing of KN1 or STM mRNA. However, when we examined mRNA levels and decay rates in dissected shoot apices or in seedlings by RT-qPCR, there were no differences in levels or stability of STM mRNA in *atrrp44a* mutants (fig. S7, A to D). We also assessed whether the trafficking function of AtRRP44A could be uncoupled from its RNA-processing activity by introducing a noncatalytic AtRRP44A mutant [AtRRP44A^{D489N}~RFP (D489N, Asp^{489}→Asn)] (*27*) driven by its native promoter into *atrrp44a* mutant trichome rescue plants. This construct fully rescued KN1 trafficking and trichome formation (fig. S7, E to G). These results suggest that AtRRP44A function in trafficking or trichome rescue is unrelated to its potential role in mRNA processing or degradation, although whether this noncanonical function

promotes RNA trafficking in the meristem remains to be tested. We next hypothesized that AtRRP44A might directly participate in KN1 or STM mRNA trafficking, for example, by recruiting their mRNA to plasmodesmata or by transporting mRNA through plasmodesmata. In support of these ideas, AtRRP44A is found in the *Arabidopsis* cell wall proteome (*35*) and the plasmodesmal proteome (*36*) despite its predominant accumulation in nuclei (Fig. 3, A, B, and D) (*16*), suggesting that a fraction of the protein may associate with plasmodesmata. We could not detect plasmodesmata localization using our native AtRRP44A-tagged lines, so to test this possibility, we modified the AtRRP44A~RFP fusion by deleting two nuclear localization signals (NLSs) and adding a nuclear export signal (NES) (AtRRP44A^{NLSΔ}~NES~RFP) to promote its export from the nucleus. Indeed, this modified protein accumulated outside of the nuclei (Fig. 3, C and D). We observed its localization
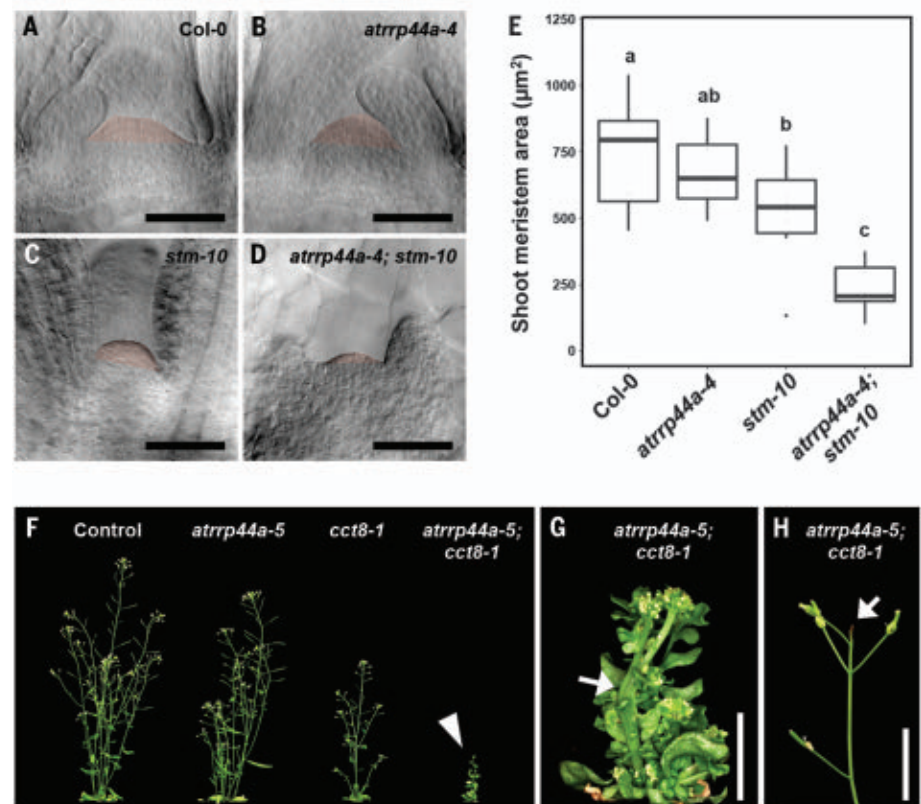

Fig. 2. *atrrp44a* mutants enhance *stm* and *cct8* phenotypes. (**A** to **E**) Meristem defects in *stm-10* were enhanced by *atrrp44a-4*. *stm-10* mutants had smaller meristems, and the meristem size was even smaller in the *atrrp44a-4; stm-10* double mutants; meristems were shaded in pink. Plants were grown for 14 days under short-day conditions. Bars topped by different letters (a, b, and c) are significantly different at $P < 0.05$ (Tukey's honest significant difference test). Scale bars, 50 μm. (**F** to **H**) *atrrp44a-5; cct8-1* double mutants developed enhanced shoot defects. *atrrp44a* and *cct8-1* single mutants had normal shoot development, although they were slightly shorter than the control lines, whereas the *atrrp44a-5; cct8-1* double mutants were severely dwarfed (arrowhead) (F), with fasciated stems (arrow) (G) and premature termination of the shoot meristem (arrow) (H). Scale bars, 1 cm.
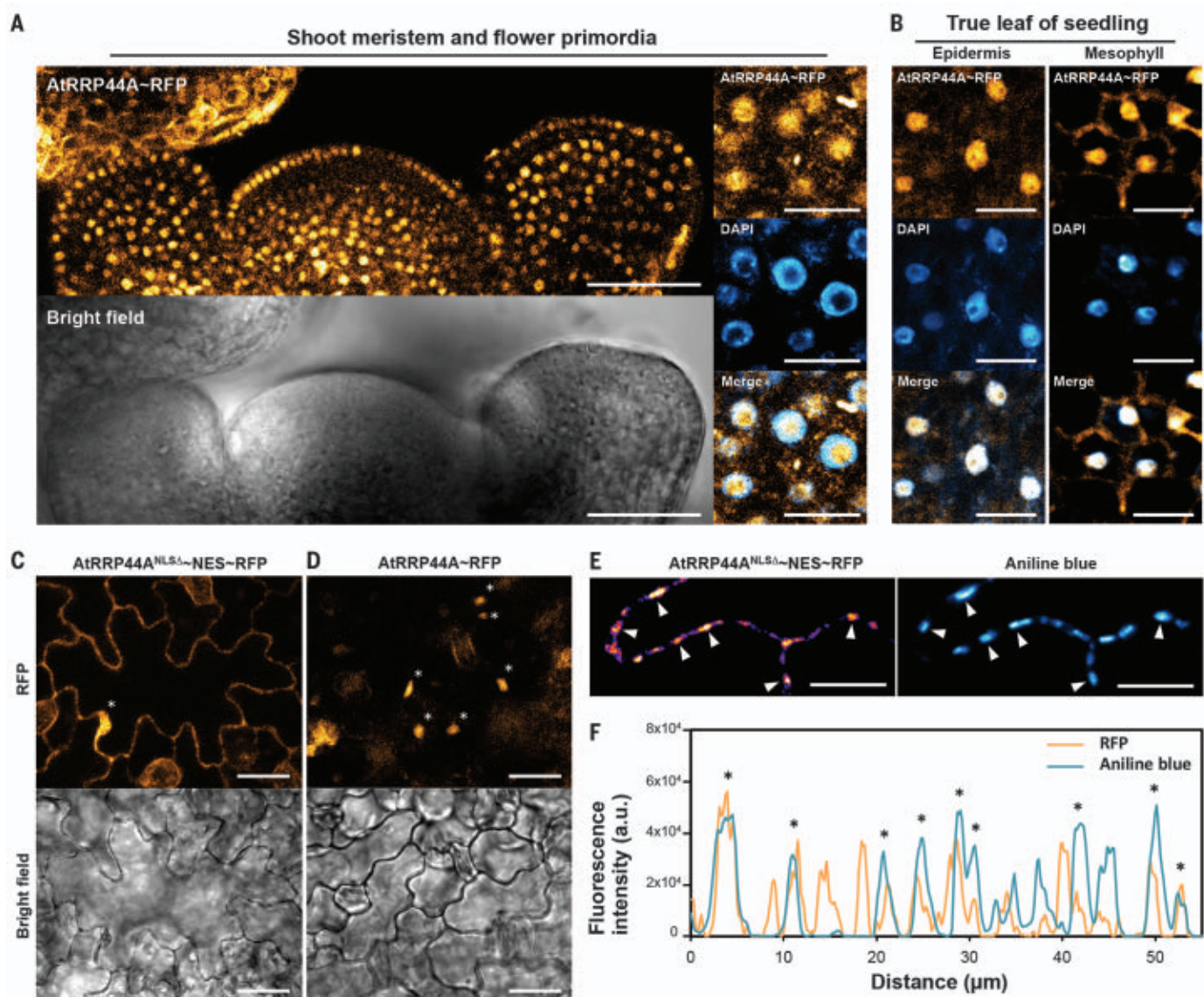
Fig. 3. Meristem expression and plasmodesmata localization of AtRRP44A.
(**A** and **B**) AtRRP44A~RFP (orange) was expressed in the shoot meristem and flower primordia (A) and epidermal and mesophyll cells in leaf primordia (B). The fusion protein localized in nuclei in all tissues, as shown by colocalization with 4′,6-diamidino-2-phenylindole (DAPI) [blue in (A) and (B)]. Scale bars, 50 μm [for the left panels in (A)] and 10 μm [for the right panels in (A) and for (B)]. (**C** and **D**) Some of the AtRRP44A^{NLSΔ}~NES~RFP protein (orange) was found outside of the nucleus and localized to the periphery of an epidermal cell (C), whereas native AtRRP44A~RFP protein (orange) was detected only in the nucleus (asterisks) (D). Scale bars, 20 μm. (**E** and **F**) AtRRP44A^{NLSΔ}~NES~RFP (pink and orange) accumulated in puncta at the boundary between epidermal cells and colocalized with aniline blue staining of plasmodesmata (arrowheads) (E). Scale bars, 10 μm. Quantitative analysis indicated that AtRRP44A accumulated in ~75% of plasmodesmata (F). Orange and blue lines show the fluorescence intensities of RFP and aniline blue along cell-cell boundaries, respectively, and asterisks represent overlaps. a.u., arbitrary units.

in puncta that colocalized with aniline blue–stained plasmodesmata (75% of plasmodesmata showed AtRRP44A^{NLSΔ}~NES~RFP signal, n = 227 of 302 plasmodesmata from 10 cells) (Fig. 3, E and F, and fig. S6C). Thus, our data support the idea that AtRRP44A interacts with plasmodesmata to promote trafficking.

AtRRP44A mediates KN1 mRNA trafficking

We next investigated whether AtRRP44A promotes mRNA trafficking by visualizing KN1 mRNA in the trichome rescue system (37). We inserted 24 repeats of the MS2 bacteriophage binding sequence (MBS) between the KN1 stop codon and 3′ untranslated region in our mesophyll-expressed trichome rescue construct to make *pRbcS::GFP~GL1~KN1^C~MBS*. This construct was introduced into plants that expressed an MS2 coat protein (MCP) ~RFP fusion in the epidermis (epiMCP) (Fig. 4A). MCP binds to MBS sequences, so if GFP~GL1~KN1^C~MBS mRNA traffics from the mesophyll to the epidermis, we expect to detect it through its binding by epiMCP~RFP in epidermal cells. Indeed, lines that were wild type (WT) for *AtRRP44A* showed many MCP~RFP puncta in the cytoplasm of epidermal cells (Fig. 4B), whereas plants that expressed MCP~RFP alone did not have puncta (fig. S8, A and B), indicating that GFP~GL1~KN1^C~MBS mRNA trafficked from mesophyll cell layers to the epidermis and formed fluorescent cytoplasmic puncta in association with MCP~RFP.
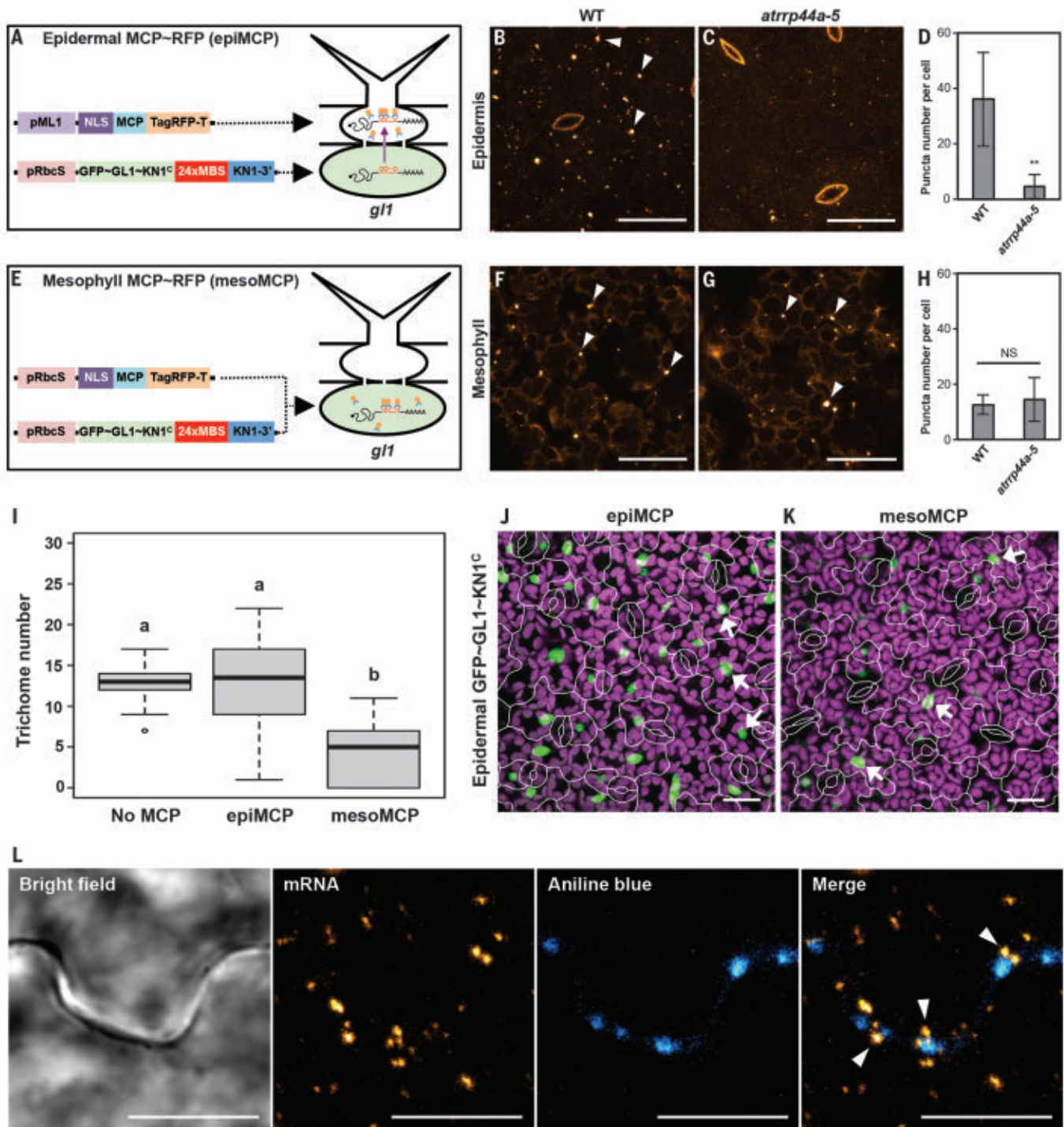
Fig. 4. AtRRP44A mediates KN1 mRNA trafficking. (**A** to **D**) To assay KN1 mRNA trafficking, GFP~GL1~KN1^C~MBS mRNA and MCP~RFP were expressed in the mesophyll layer and epidermis, respectively (epiMCP) (A). In lines WT for *AtRRP44A*, KN1 mRNA trafficked to the epidermis and formed many speckles in the cytoplasm of epidermal cells (orange, arrowheads) (B), whereas trafficking was strongly suppressed in *atrrp44a-5* mutants, which had approximately eightfold fewer epidermal fluorescent puncta compared with WT [(C) and (D)]. Scale bars, 20 μm. **$P < 0.01 (Welch's *t* test). (**E** to **H**) Coexpression of GFP~GL1~KN1^C~MBS mRNA and mesoMCP~RFP (E) indicated that *atrrp44a-5* mutants had similar numbers of KN1 mRNA puncta in mesophyll cells as in WT [orange, arrowheads in (F) to (H); *P* = 0.145, Welch's *t* test]. Scale bars, 20 μm.

Error bars in (D) and (H) denote SD. NS, nonsignificance. (**I** to **K**) Expression of mesoMCP~RFP significantly reduced trichome rescue compared with epiMCP or lines without MBS–MCP expression (no MCP) (I) and also decreased GFP~GL1~KN1^C accumulation in the epidermis (K) in comparison with lines expressing epiMCP~RFP (J). Bars topped by different letters (a and b) are significantly different at *P* < 0.01 (Tukey's honest significant difference test). Arrows point to representative nuclei with GFP~GL1~KN1^C fusion protein. Chlorophyll autofluorescence is shown in magenta. White lines show the outlines of epidermal cells. Scale bars, 20 μm. (**L**) GFP~GL1~KN1^C~MBS mRNA (orange, arrowheads) associated with plasmodesmata, as shown by colocalization with aniline blue signals (blue). Scale bar, 10 μm.

By contrast, when we introduced the MBS–MCP system into *atrrp44a* mutants, their epidermal cells had approximately eightfold fewer fluorescent puncta (Fig. 4, C and D) despite similar levels of GFP~GL1~KN1^C protein in mesophyll cells (fig. S8, C and D). The puncta in *atrrp44a-5* were also smaller, suggesting that they contained fewer GFP~GL1~KN1^C~MBS mRNA molecules (fig. S8E). To confirm that this apparent loss of mRNA trafficking was not due to a reduction of GFP~GL1~KN1^C~MBS mRNA in *atrrp44a* mutants, we expressed MCP~RFP in mesophyll cell layers (mesoMCP, Fig. 4E). As expected, the mutants had similar numbers of KN1 mRNA puncta in mesophyll cells as the WT lines (Fig. 4, F to H). Thus, our data suggest that AtRRP44A is required for the cell-to-cell trafficking of KN1 mRNA in the trichome system, although its contribution to RNA trafficking in the meristem remains to be tested. To support this role, we examined whether AtRRP44A could bind to STM mRNA in vivo, using RNA immunoprecipitation–qPCR from extracts of shoot apices of *pAtRRP44A:: AtRRP44A~RFP; atrrp44a-1* plants. Indeed, we found that AtRRP44A was associated with STM transcripts in vivo (fig. S8F).

KN1 mRNA trafficking promotes protein trafficking

We next investigated whether KN1 mRNA trafficking was important in the selective transport of a KN1 signal, by scoring trichome rescue in the AtRRP44A WT lines expressing mesoMCP~RFP (Fig. 4E). We reasoned that binding of MCP~RFP to GFP~GL1~KN1^C~MBS mRNA in mesophyll cells might inhibit its trafficking because of the large size of the mRNP complex, which is similar to a previous study (*38*). Trichome rescue was reduced in mesoMCP lines compared to epiMCP lines or our original trichome rescue lines without the MBS–MCP system (no MCP) (Fig. 4I). GFP~GL1~KN1^C accumulation in the epidermis was also consistently decreased in mesoMCP lines compared with that of epiMCP lines (*n* = 16 plants from two independent transgenic lines) (Fig. 4, J and K). This result suggests that the cell-to-cell trafficking of KN1 mRNA promotes protein trafficking, for example, as an mRNP complex, and/or that the transported mRNA is translated de novo in the epidermis. Other homeodomain proteins also have the capacity to bind their mRNAs (*39*). Therefore, our data support the idea that AtRRP44A-mediated transport of a homeodomain mRNA plays a role in the cell-to-cell trafficking of a KN1 signal, possibly as an mRNP complex.

KN1 mRNA transiently associates with plasmodesmata

To support the idea that KN1 mRNA traffics cell to cell, we investigated whether KN1 mRNA associates with plasmodesmata, like other mobile mRNAs in plants (*40*, *41*). By imaging the interface between epidermal cells, we found colocalization of GFP~GL1~KN1^C~MBS mRNA with aniline blue–stained plasmodesmata (Fig. 4L). By using time-lapse imaging, we found that KN1 mRNA puncta moved freely through the cytoplasm until they "met" a plasmodesma, where they paused for ~1 to 6 min before leaving (movies S1 and S2). We also observed KN1 mRNA–plasmodesmata interactions at the interface of epidermal and mesophyll cells (movie S3) and between leaf epidermal cells when expressed in a heterologous tobacco system (movie S4). Thus, our data support the idea that KN1 mRNA associates with plasmodesmata to traffic from cell to cell.

We demonstrate a role of KN1 mRNA trafficking in the selective transport of a KN1 signal and the role of AtRRP44A in this process (fig. S9). This role of AtRRP44A appears to be independent of its function in RNA metabolism and may involve additional factors that prevent its RNA-processing activity, despite its binding to the mobile homeodomain mRNA. Our finding that trafficking of maize KN1 mRNA is mediated by AtRRP44A in *Arabidopsis* suggests that this mechanism is required for plant stem cell function in diverse plants. Whether RRP44A functions in transport of other plant mobile RNAs remains to be seen.

REFERENCES AND NOTES

1. Z. P. Li, A. Paterlini, M. Glavier, E. M. Bayer, *Cell. Mol. Life Sci.* **78**, 799–816 (2021).
2. M. Kitagawa, D. Jackson, *Plants* **6**, 12 (2017).
3. J. O. Brunkard, P. C. Zambryski, *Curr. Opin. Plant Biol.* **35**, 76–83 (2017).
4. W. J. Lucas et al., *Science* **270**, 1980–1983 (1995).
5. L. Liu et al., *Cell Rep.* **23**, 1879–1890 (2018).
6. N. Winter, G. Kollwig, S. Zhang, F. Kragler, *Plant Cell* **19**, 3001–3018 (2007).
7. X. M. Xu et al., *Science* **333**, 1141–1144 (2011).
8. L. Yang et al., *Curr. Biol.* **29**, 2465–2476.e5 (2019).
9. T. H. Ghate, P. Sharma, K. R. Kondhare, D. J. Hannapel, A. K. Banerjee, *Plant Mol. Biol.* **93**, 563–578 (2017).
10. K. J. Lu, N. C. Huang, Y. S. Liu, C. A. Lu, T. S. Yu, *RNA Biol.* **9**, 653–662 (2012).
11. M. Notaguchi, S. Wolf, W. J. Lucas, *J. Integr. Plant Biol.* **54**, 760–772 (2012).
12. G. Kim, M. L. LeBlanc, E. K. Wafula, C. W. dePamphilis, J. H. Westwood, *Science* **345**, 808–811 (2014).
13. J. Y. Kim, Y. Rim, J. Wang, D. Jackson, *Genes Dev.* **19**, 788–793 (2005).
14. D. G. Oppenheimer, P. L. Herman, S. Sivakumaran, J. Esch, M. D. Marks, *Cell* **67**, 483–493 (1991).
15. R. Fekih et al., *PLOS ONE* **8**, e68529 (2013).
16. W. Zhang, C. Murphy, L. E. Sieburth, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15981–15985 (2010).
17. N. Kumakura, H. Otsuki, M. Tsuzuki, A. Takeda, Y. Watanabe, *PLOS ONE* **8**, e79219 (2013).
18. H. Ohkura et al., *EMBO J.* **7**, 1465–1473 (1988).
19. M. J. Snee et al., *Genetics* **203**, 749–762 (2016).
20. R. Tomecki et al., *Nucleic Acids Res.* **42**, 1270–1290 (2014).
21. P. Mitchell, E. Petfalski, A. Shevchenko, M. Mann, D. Tollervey, *Cell* **91**, 457–466 (1997).
22. C. Kilchert, "RNA exosomes and their cofactors" in *The Eukaryotic RNA Exosome*, J. LaCava, Š. Vaňáčová, Eds. (Springer, 2020), pp. 215–235.
23. J. C. Zinder, C. D. Lima, *Genes Dev.* **31**, 88–100 (2017).
24. C. Kilchert, S. Wittmann, L. Vasiljeva, *Nat. Rev. Mol. Cell Biol.* **17**, 227–239 (2016).
25. A. Dziembowski, E. Lorentzen, E. Conti, B. Séraphin, *Nat. Struct. Mol. Biol.* **14**, 15–22 (2007).
26. E. Lorentzen, J. Basquin, R. Tomecki, A. Dziembowski, E. Conti, *Mol. Cell* **29**, 717–728 (2008).
27. N. Kumakura et al., *Plant Biotechnol. (Tsukuba)* **33**, 77–85 (2016).
28. D. L. Makino, M. Baumgärtner, E. Conti, *Nature* **495**, 70–75 (2013).
29. J. Y. Kim, Z. Yuan, D. Jackson, *Development* **130**, 4351–4362 (2003).
30. R. Balkunde, M. Kitagawa, X. M. Xu, J. Wang, D. Jackson, *Plant J.* **90**, 435–446 (2017).
31. S. Kanrar, O. Onguka, H. M. Smith, *Planta* **224**, 1163–1173 (2006).
32. J. A. Long, E. I. Moan, J. I. Medford, M. K. Barton, *Nature* **379**, 66–69 (1996).
33. H. Lange et al., *PLOS Genet.* **10**, e1004564 (2014).
34. J. A. Chekanova et al., *Cell* **131**, 1340–1353 (2007).
35. E. M. Bayer et al., *Proteomics* **6**, 301–311 (2006).
36. L. Fernandez-Calvino et al., *PLOS ONE* **6**, e18880 (2011).
37. T. Lionnet et al., *Nat. Methods* **8**, 165–170 (2011).
38. G. Haimovich et al., *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9873–E9882 (2017).
39. J. Dubnau, G. Struhl, *Nature* **379**, 694–699 (1996).
40. A. Sambade et al., *Traffic* **9**, 2073–2088 (2008).
41. K. R. Luo, N. C. Huang, T. S. Yu, *Plant Physiol.* **177**, 604–614 (2018).

CORONAVIRUS

Heterologous infection and vaccination shapes immunity against SARS-CoV-2 variants

Catherine J. Reynolds[1]†, Joseph M. Gibbons[2]†, Corinna Pade[2]†, Kai-Min Lin[1], Diana Muñoz Sandoval[1], Franziska Pieper[1], David K. Butler[1], Siyi Liu[1], Ashley D. Otter[3], George Joy[4], Katia Menacho[4], Marianna Fontana[5], Angelique Smit[5], Beatrix Kele[4], Teresa Cutino-Moguel[4], Mala K. Maini[6], Mahdad Noursadeghi[6], COVIDsortium Immune Correlates Network‡, Tim Brooks[3], Amanda Semper[3], Charlotte Manisty[4,7], Thomas A. Treibel[4,7], James C. Moon[4,7], COVIDsortium Investigators‡, Áine McKnight[2]§, Daniel M. Altmann[8]§, Rosemary J. Boyton[1,9]*§

The impact of the initial severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infecting strain on downstream immunity to heterologous variants of concern (VOCs) is unknown. Studying a longitudinal healthcare worker cohort, we found that after three antigen exposures (infection plus two vaccine doses), S1 antibody, memory B cells, and heterologous neutralization of B.1.351, P.1, and B.1.617.2 plateaued, whereas B.1.1.7 neutralization and spike T cell responses increased. Serology using the Wuhan Hu-1 spike receptor binding domain poorly predicted neutralizing immunity against VOCs. Neutralization potency against VOCs changed with heterologous virus encounter and number of antigen exposures. Neutralization potency fell differentially depending on targeted VOCs over the 5 months from the second vaccine dose. Heterologous combinations of spike encountered during infection and vaccination shape subsequent cross-protection against VOC, with implications for future-proof next-generation vaccines.

After experiencing >18 months of the COVID-19 pandemic, immunological analysis has shifted to issues of response durability, boosting, and mitigation against future variants of concern (VOCs). Researchers thus are confronting viral and population heterogeneities that affect immunity and future protection (1–4). Individuals are either immunologically naïve to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or have been infected with either the original Wuhan Hu-1 strain or one of the Alpha to Delta VOCs, which we will refer to by full Pango lineage terms. Infection-naïve individuals, those infected by ancestral SARS-CoV-2 Wuhan Hu-1 or a VOC, may have received different numbers of vaccine doses. Thus, there is a spectrum of individuals spanning from the immunologically naïve to those who have experienced one, two, three exposures (and four with boosting) to homologous or heterologous spike sequences. The challenge is to understand whether different antigen exposure combinations are associated with the same

quality, quantity, and durability of immunity and ability to cross-protect against other VOCs.

Using a cohort of UK healthcare workers (HCWs) for whom we have extensive, longitudinal, clinical, transcriptomic, and immunologic characterization (5–10), we address these questions by immunological comparison of BNT162b2 (Pfizer-BioNTech vaccine) vaccinees, with or without infection. We and others have shown the boosting effect of prior infection by the Wuhan Hu-1 strain on the response to vaccination with homologous spike (8, 9, 11–14), and others are decoding changes in immune programming between the first and second dose (15). We compared the impact of prior infection with Wuhan Hu-1 or B.1.1.7. (Alpha VOC) in the context of vaccination on T and memory B cell (MBC) responses, cross-neutralization against VOCs, durability of immunity, and susceptibility to B.1.617.2 (Delta VOC) breakthrough infection. We considered the extent to which first encounter with the spike sequence shapes subsequent response features to explore the possible effect of immune imprinting or "original antigenic sin" (16–18).

B and T cell immunity after three homologous antigen exposures

We analyzed this longitudinal vaccine cohort (n = 51) at 20 d [interquartile range (IQR) = 7] after receiving a second BNT162b2 dose (fig. S1 and table S1). Twenty-five HCWs were infected with SARS-CoV-2 approximately synchronously and coincidentally with peak transmission in London in March 2020 (6). At 1 year, nucleoprotein (N) antibody (Ab) responses had waned but were still positive (anti-N Ab levels expressed as a cutoff index of ≥1.0 were classified as positive), 16% having fallen to sub-

threshold levels (Fig. 1A). Two (2/25, 8%) HCWs appeared to have been reinfected (increase in N) and two (2/26, 8%) newly infected HCWs (positive N having been previously negative) were among the previously uninfected cohort (Fig. 1A). T cell responses to N were sustained at 1 year in most previously infected HCWs (22/24, 92%) (fig. S2). All previously infected HCWs with positive anti–S1 receptor binding domain (RBD) responses at 16 to 18 weeks had sustained responses at 28 to 30 weeks (22/24, 92%), which substantially increased after single-dose vaccination (36-fold increase in Ab titer) (Fig. 1B). This response plateaued (i.e., only increased a further 1.4-fold) after second-dose vaccination (third antigen exposure). The previously infected HCWs (2/24, 8%) with unrecordable anti-S1 RBD responses at 16 to 18 and 28 to 30 weeks showed a 155-fold increase in anti-S1 response after second-dose vaccination, demonstrating the benefit of a third antigen exposure in poor Ab responders (19, 20). Infection-naïve HCWs had incrementally increased Ab responses after first- and second-dose vaccination, achieving about half the Ab titer of their previously infected counterparts 20 days after their second vaccine dose (Fig. 1B).

T cell responses to the spike mapped epitope peptide (MEP) pool (table S2) increased after second-dose vaccination in previously infected individuals (i.e., third exposure to antigen) (P = 0.0294; Fig. 1C). T cell responses after second-dose vaccination in infection-naïve HCWs achieved similar levels to those seen at 16 to 18 weeks after natural SARS-CoV-2 infection. However, there were fewer nonresponders (2/23, 9%) in double-vaccinated infection-naïve individuals compared with (5/25, 20%) 16 to 18 weeks after natural infection.

Infection-naïve individuals showed incremental increases in neutralizing Ab (nAb) to authentic Wuhan Hu-1 virus and VOCs after the first and second vaccination, with the largest fold increase occurring after the second vaccine dose (Wuhan Hu-1,123-fold; B.1.1.7, 589-fold; B.1.351, 445-fold; P.1, 1765-fold; and B.1.617.2, 39-fold). There was a wider heterogeneity of nAb response to B.1.351, P.1, and B.1.617.2 VOCs compared with B.1.1.7 and Wuhan Hu-1 (Fig. 1D). Individuals who had experienced SARS-CoV-2 infection before vaccination also showed the largest fold nAb increase on second antigen exposure (i.e., after the first vaccine dose) (Wuhan Hu-1, 27-fold; B.1.1.7, 52-fold; B.1.351, 65-fold; P.1, 63-fold; and B.1.617.2, 21-fold) (Fig. 1E). By contrast, nAb responses against Wuhan Hu-1 and VOCs B.1.351, P.1, and B.1.617.2 plateaued or decreased between the second and third antigen exposure. This was not the case for nAb responses against Wuhan Hu-1 and B.1.1.7, which increased sixfold and 23-fold, respectively, between the second and third antigen exposure (Fig. 1E).

[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]Royal Free London NHS Foundation Trust, London, UK. [6]Division of Infection and Immunity, University College London, London, UK. [7]Institute of Cardiovascular Science, University College London, London, UK. [8]Department of Immunology and Inflammation, Imperial College London, London, UK. [9]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 and are co-first authors.
‡The members of the COVIDsortium Immune Correlates Network and COVIDsortium Investigators are listed in the supplementary materials.
§These authors contributed equally to this work and are co-senior authors.

The potency of nAb response after the third antigen exposure varied depending on the VOC being neutralized. Previously infected HCWs who had a low nAb IC$_{50}$ after the first vaccine dose caught up after the second vaccine dose (i.e., the third antigen exposure). Heterologous neutralizing IC$_{50}$s were lower than those against homologous Wuhan Hu-1, and were at risk of falling below a threshold for protection as levels waned (*21*, *22*) (1F, S3).

There was a positive correlation between S1 RBD-binding Ab and neutralization (IC$_{50}$), with previously infected individuals showing higher nAb IC$_{50}$ and S1 RBD binding (Fig. 1G). Wuhan-Hu-1 sequence–specific S1 RBD binding correlated less well with nAb IC$_{50}$ against VOCs B.1.351, P.1, and B.1.617.2, especially for infection-naïve, double-vaccinated HCWs. To establish whether this was due to different sequences in the VOC RBD, the relationship between VOC RBD-binding titers and neutralization was explored. There was positive correlation between the Roche S1 RBD and VOC RBD specific binding (fig. S4). The weaker correlations between VOC nAb titers and VOC RBD binding indicate that antibodies targeting regions outside of the RBD may contribute to neutralization, so the binding titer is not predictive of neutralization (Fig. 1H). For example, a two-dose-vaccinated HCW with an S1 RBD (Wuhan Hu-1)–binding titer of 2950 U/ml, considered strongly positive, showed divergent neutralization of VOCs, with IC$_{50}$s of Wuhan Hu-1, 6,071, B.1.1.7, 1,037; B.1.351, 301; P.1, 560, and B.1.617.2 zero 20 d after second-dose vaccination that all fell to undetectable levels 18 weeks later. This individual was later infected by B.1.617.2. Thus, S1 serology data using the Wuhan Hu-1 S1 RBD and VOC sequence is an unreliable marker for neutralization potency against VOCs.

For infection-naïve subjects, the frequency of S1-specific MBC 20 d after first-dose vaccination was lower compared with 16 to 18 weeks after infection. Conversely, MBCs were amplified in previously infected HCWs to a level similar to that seen after two doses of vaccine (Fig. 1I). As expected from the S1 RBD-binding Ab and nAb responses, MBC responses plateaued, with no further enhancement after a third antigen exposure (i.e., double-vaccinated previously infected HCWs) (Fig. 1, I and J). For infection-naïve HCWs, the MBC frequency for specificity to S1 containing the N501Y, E484K, and K417N B.1.351 mutations was lower after one vaccine dose (Fig. 1K, left panel). After two (two-dose vaccination or single-dose vaccination and prior infection) or three (two-dose vaccination and prior infection) antigen exposures, the MBC response was maintained whether stimulated by S1 or S1-containing mutations found in B.1.351 and B.1.617.2 (Fig. 1K, middle and right panels). There was a high frequency of MBCs able to recognize S1 and

S1-containing VOC-specific mutations, but this did not always correlate well with a VOC-specific nAb response (fig. S5). Authentic P.1 (*r* = 0.529, *P* = 0.0138) and B.1.617.2 (*r* = 0.548, *P* = 0.0102) nAb responses of double-vaccinated HCWs with and without prior Wuhan Hu-1 infection positively correlated with the MBC frequency against S1 containing the B.1.617.2–specific mutations (fig. S5C).

Next, we examined the T cell response against VOCs B.1.1.7, B.1.351, P.1, and B.1.617.2 peptide pools and Wuhan Hu-1 matched pools and individual peptides with substituted epitopes covering N501Y and D1118H (table S2 and Fig. 2, A and B). T cell responses in doubly vaccinated infection-naïve individuals were variable. Previously infected double-vaccinated individuals showed a significantly increased T cell response to B.1.1.7 (*P* = 0.0153) and B.1.617.2 (*P* = 0.0283) peptide pools and N501Y (*P* = 0.0156) variant peptide compared with the respective Wuhan Hu-1 pools and peptide (Fig. 2B). A cumulative increase in T cell response to the Wuhan Hu-1 spike MEP pool was observed as the number of antigen exposures increased (one to two exposures, *P* = 0.0003; two to three exposures, *P* = 0.0294; Fig. 2C). There was a similarly increased T cell response to the B.1.1.7 variant peptide pool (one to two exposures, *P* = 0.0008; two to three exposures, *P* = 0.0004; Fig. 2E) depending on number of antigen exposures, indicating a heteroclitic response to variant peptides (*23*) (Fig. 2, D and E).

VOC mutation alters T cell effector program

Immunization of transgenic mice expressing human HLADRB1*0401 were used to determine whether any of the variant peptides behave as an altered peptide ligand, eliciting a differential T cell program (*9*). In silico analysis of P.1 variant peptides indicated that most are not predicted to bind to common HLAII alleles found in the UK population (table S3). The N501Y mutation is predicted to change from strong to weak binding to HLADRB1*0401. Interferon γ (IFNγ) T cell responses in mice primed with Wuhan-Hu-1 peptide pool were ablated for the 501Y variant peptide (found in the B.1.1.7, B.1.351, and P.1 VOCs (Fig. 2F) when presented with HLA-DRB1*0401 (*n* = 7, *P* = 0.006). Altered CD4 recognition of the N501Y mutation merits further attention because it is a frequent replacement among reported sequences, comprising a convergent mutational signature shared by the Alpha, Beta, and Gamma lineages (*24–26*). Thus far, the lack of T cell responses to 501Y has been described in terms of absence of CD8 recognition (*27*). Our transcriptomic analysis showed an absence of immune effector response to the variant peptide, although induction of the regulatory T cell transcription factor FOXP3 did occur, indicating that the variant peptide

may exert an altered peptide ligand effect on T cell function, switching from effector to regulatory (Fig. 2G). This was further confirmed by quantitative polymerase chain reaction (qPCR) for IRF4 and by CD4 expression of FoxP3 protein, indicating that mutant epitopes may subvert T cell activation into a regulatory program (Fig. 2, G to I).

Mice were primed with either B.1.617.2 or Wuhan Hu-1 matched specific peptide pools. In each case, T cells responded on challenge to the pool with which they were primed and not to the heterologous panel (*P* = 0.0022; Fig. 2J). That is, the T cell repertoire distinguishes VOC mutations, specifically recognizing the B.1.617.2 epitopes. In silico analysis of the B.1.617.2 variant peptides indicated that most are predicted to bind one or more of the common HLAII alleles found in the UK population (table S4). The D950N mutation is predicted to change from HLADRB1*0401 binding from weak to strong. T cell responses after priming with B.1.617.2 peptide pool were present for the D950N variant peptide but not after priming with the Wuhan Hu-1 pool (*P* = 0.0022; Fig. 2J).

Heterologous B.1.1.7 infection shapes subsequent immunity

We then looked at the impact of infection with B.1.1.7 during the second wave on immune responses in individuals given vaccine expressing Wuhan Hu-1 spike. For this part of the study, an additional 358 HCWs were followed up at 55 to 57 weeks after recruitment into the study (fig. S1). Of these, 63 had been infected with SARS-CoV-2 during the first Wuhan Hu-1 wave. Fifty-three of previously uninfected HCWs (53/296, 18%) were identified by virtue of longitudinal N serology (from baseline to 16 to 18, 28 to 30, 42, 54, and 55 to 57 weeks) as having been newly infected during the B.1.1.7 s wave. Thirty-six (68%) were double-vaccinated and eight (15%) unvaccinated at the time of follow-up (fig. S1 and tables S5 and S6). Five previously infected HCWs were reinfected (5/63, 8%) during the second wave. S1 IgG titers were measured at 55 to 57 weeks (fig. S6B). The median date of a positive SARS-CoV-2 PCR test was 28 December 2020 (IQR = 22 d). The B.1.1.7 VOC accounted for 94.7% of SARS-CoV-2 infections in central London in the 2 weeks leading up to 2 January 2021 (*28*); we thus made the starting assumption that for most HCW-infected individuals during the second wave, the infecting strain was B.1.1.7. To support this, we compared the Wuhan Hu-1 and B.1.1.7 nAb IC$_{50}$ results for unvaccinated (*n* = 24 and 8), single-dose vaccinated (*n* = 25 and 9), and two-dose vaccinated (*n* = 24 and 34) HCWs infected during the first and second UK waves by Wuhan Hu-1 and B.1.1.7, respectively (fig. S7). In unvaccinated HCWs, those infected during the first wave (fig. S7A, LHS) had nAb
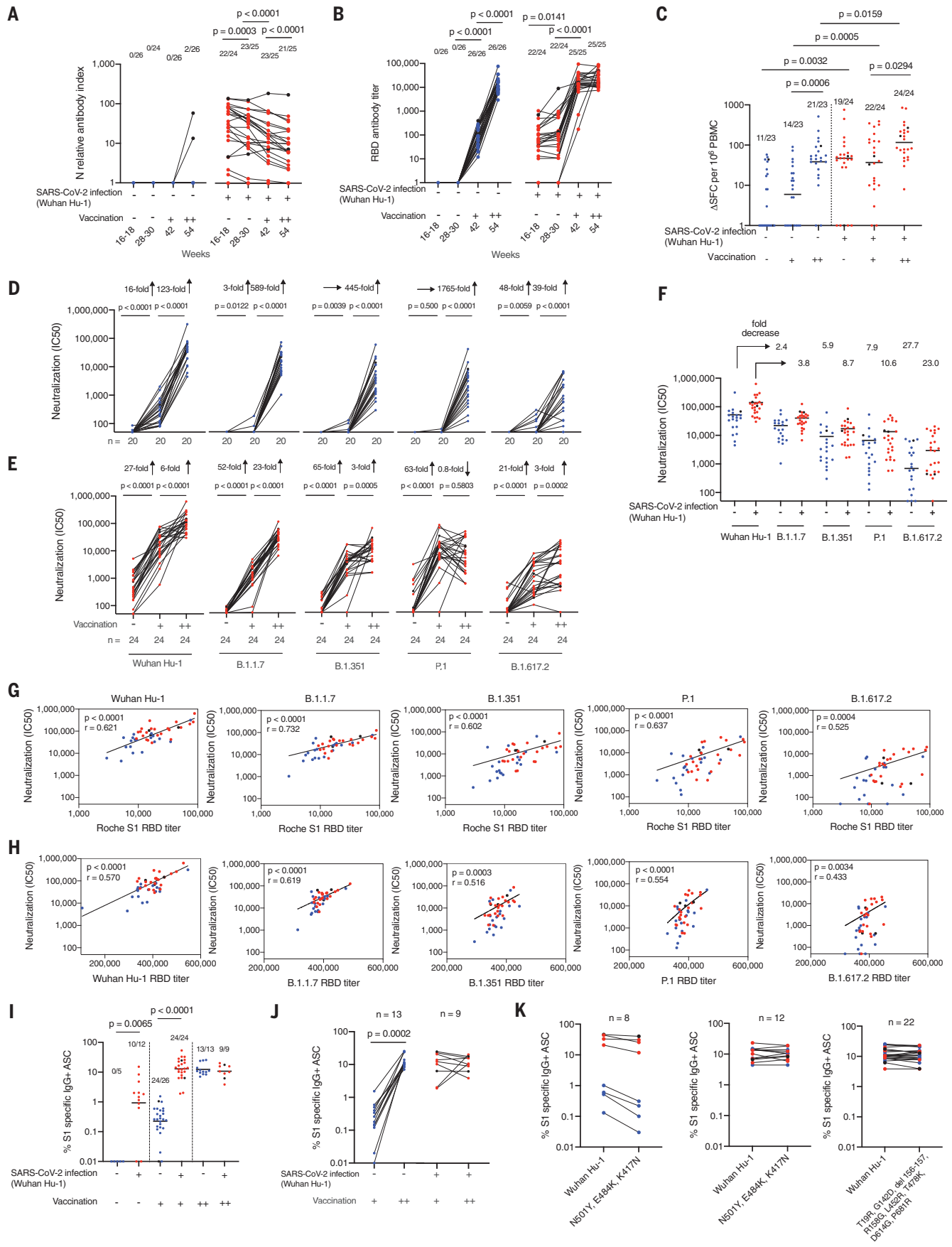
Fig. 1. T and B cell immunity to infection with Wuhan Hu-1 SARS-CoV-2 and after one and two doses of BNT162b2 vaccine in infection naïve and previously infected HCWs. (**A**) N Ab and (**B**) S1 RBD serum Ab titers measured by ECLIA in HCWs with (red, $n = 25$) and without (blue, $n = 26$) laboratory-confirmed SARS-CoV-2 infection at 16 to 18 weeks after the start of the first UK epidemic wave and 2 to 3 weeks after first (+, 42 weeks) and second (++, 54 weeks) dose of BNT162b2 vaccine. In all figures, HCWs without laboratory-confirmed infection are shown in blue and HCWs with laboratory-confirmed infection are shown in red. HCWs with new infection or reinfection after 30 weeks are shown in black. (**C**) Magnitude of T cell responses to spike MEP pool in HCWs with (red, $n = 24$) and without (blue, $n = 23$) laboratory-confirmed SARS-CoV-2 infection 16 to 18 weeks after the start of the first UK epidemic wave and 2 to 3 weeks after first (+, 42 weeks) and second (++, 54 weeks) dose of vaccine. Neutralizing Ab titer (IC_{50}) against authentic Wuhan Hu-1 live virus and B.1.1.7, B.1.351, P.1, and B.1.617.2 VOCs is plotted longitudinally for (**D**) SARS-CoV-2 infection naïve (blue, $n = 20$) and (**E**) previously Wuhan Hu-1 infected HCWs (red, $n = 24$) unvaccinated and after the first and second dose of BNT162b2 vaccine. (**F**) IC_{50} nAb against Wuhan Hu-1 and VOCs after two-dose BNT162b2 vaccination in previously infected HCWs (red, $n = 24$) and infection-naïve HCWs (blue, $n = 20$). (**G**) Correlation between Roche S1 RBD Ab titer and nAb (IC_{50}) against SARS-CoV-2 authentic Wuhan Hu-1 live virus and the B.1.1.7, B.1.351, P.1, and B.1.617.2 VOCs. (**H**) Correlation between Wuhan Hu-1,

VOC RBD, and corresponding live virus in two-dose BNT162b2 vaccinated previously infected (red, $n = 24$) and SARS-CoV-2 infection-naïve (blue, $n = 20$) HCWs. (**I**) Percentage of S1-specific IgG$^+$ antibody-secreting cells (ASCs) in SARS-CoV-2 infection naïve and previously infected HCWs at 16 to 18 weeks and after first- and second-dose BNT162b2 vaccination. (**J**) Percentage of S1-specific IgG$^+$ ASCs plotted pairwise after first- and second-dose vaccination for HCWs with (red, $n = 9$) and without (blue, $n = 13$) prior SARS-CoV-2 infection. (**K**) Left, percentage of IgG$^+$ ASCs specific for Wuhan Hu-1 S1 protein or spike protein containing the mutations N501Y, E484K, and K417N in HCWs with (red) and without (blue) prior SARS-CoV-2 infection after first-dose vaccination ($n = 4$ per group). Middle and right panels, percentage of IgG$^+$ ASCs specific for Wuhan Hu-1 S1 or S1 spike protein containing N501Y, E484K, and K417N (B.1.351 VOC) mutations or S1 spike protein containing T19R, G142D, del 156-157, R158G, L452R, T478K, D614G, and P681R (B.1.617.2 VOC) mutations after second-dose BNT162b2 vaccination (middle panel, $n = 6$ per group; right panel, infection naïve HCWs are indicated in blue, $n = 13$, and previously Wuhan Hu-1 infected HCWs are indicated in red, $n = 9$). In (A), (B), (D), (E), and (H), Wilcoxon matched-pairs signed rank test was used; in (C) and (I), Mann-Whitney U test was used; and in (G), Spearman's rank correlation was used. New infection (in the infection-naïve group, $n = 2$) and reinfection (in the previously infected group, $n = 2$) data points shown in black were excluded from statistical analysis. SFCs, spot-forming cells.

IC_{50} responses against Wuhan Hu-1, but negligible or no nAb IC_{50} response against B.1.1.7 ($P < 0.0001$). By contrast, HCWs infected as B.1.1.7 peaked (fig. S7A, RHS) had similar nAb IC_{50}s against B.1.1.7 and Wuhan Hu-1 live virus. In Wuhan Hu-1–infected HCWs, the Wuhan Hu-1 and B.1.1.7 nAb IC_{50} increased and remained significantly different after one-dose vaccination ($P < 0.0001$) and two-dose vaccination ($P < 0.0001$), but the differential between them decreased (Fig. 3F and fig. S7, B and C).

We set out to investigate the impact of heterologous antigen exposure with the B.1.1.7 variant. A confounder in assessing comparative postvaccination immunity between HCWs infected during the first wave (Wuhan-Hu-1) and those infected during the second wave (B.1.1.7) is that they differ with respect to the infecting strain and in the time interval between infection and vaccination. We hypothesized that observed differences in immunity may relate to recall differences between heterologous spike sequences [immune imprinting (16–18)] and/or to differences in affinity maturation of the Ab repertoire with time from initial infection. Somatic hypermutation and breadth of neutralizing response is known to improve with time from infection (29–31), but there may also be intrinsic differences between the breadth and hierarchy of neutralizing responses primed by different variants (32). We initially compared nAb profiles for neutralization of Wuhan Hu-1 and VOC Alpha to Delta at a similar number of weeks after PCR-confirmed SARS-CoV-2 infection with Wuhan Hu-1 during the first wave and B.1.1.7 during the second wave (all unvaccinated HCWs). These data support the contention that there are qualitatively and quantitatively differential patterns, even matching for time point (fig. S9).

After a single vaccine dose, infection-naïve HCWs had an S1 RBD response similar to that seen after infection with B.1.1.7. Those HCWs who were infected by Wuhan Hu-1 12 months earlier had a 3.9-fold higher S1-RBD response than those infected more recently by the B.1.1.7 variant ($P = 0.004$; Fig. 3A). After two vaccine doses, HCWs infected during March 2020 by Wuhan Hu-1 ($n = 70$), having had three homologous antigen exposures, had a 1.5-fold higher S1-RBD response than those encountering three antigen exposures, with one of these being heterologous B.1.1.7 ($n = 36$, $P = 0.0188$; Fig. 3A). The B.1.1.7–infected two-dose vaccinated HCWs had an S1-RBD response similar to that of the two-dose infection-naïve vaccinees. Infection-naïve vaccinees had a 1.8-fold lower response compared with the Wuhan Hu-1 previously infected group ($n = 241$, $P < 0.0001$; Fig. 3A). The implication of this is that the phenomenon of enhanced vaccine responses by infection, which has been reproducibly described by us and others (8, 9, 11–14), is less effective if the infection involves heterologous spike from a VOC.

After two vaccine doses, individuals infected during the Wuhan Hu-1 wave had a higher T cell response to the B.1.1.7 peptide pool than those infected during the B.1.1.7 wave ($P = 0.0412$), arguing for stronger T cell priming in those with three homologous exposures (Fig. 3, B to D).

We next considered the effect of B.1.1.7 infection in single and double vaccination. In two-dose vaccinated individuals, B.1.1.7 infection resulted in a different hierarchy of nAb IC_{50} responses against VOCs compared with after Wuhan Hu-1 infection: Wuhan Hu-1 > B.1.1.7 > B.1.351 > P.1 > B.1.617.2. In the context of heterologous B.1.1.7 infection and two-dose vaccination, the hierarchy changed to B.1.1.7 >

Wuhan Hu-1 > B.1.617.2 > P.1 > B.1.351 (Fig. 3, E and F). This indicates a process of selective discriminative heterologous imprinting after exposure to B.1.1.7 infection (Fig. 3, E and F, and fig. S8). As predicted for immune imprinting by SARS-CoV-2 variants, the first encounter imparts a differential subsequent pattern (16–18).

To better understand the relationship between differential nAb IC_{50} and S1 RBD binding, we correlated the two. First we compared the impact of two homologous antigen exposures [either two-dose vaccination (blue, infection-naïve) or prior Wuhan Hu-1 infection followed by single-dose vaccination (red, Wuhan Hu-1)] with heterologous exposure through B.1.1.7 infection [single-dose vaccination and B.1.1.7 infection (green, B.1.1.7)]. Two-dose vaccinated infection-naïve HCWs (blue) achieved higher neutralization against authentic Wuhan Hu-1 live virus than those with heterologous infection by B.1.1.7 and single-dose vaccination (green) ($P = 0.0036$; Fig. 3G and fig. S10A). Individuals infected by B.1.1.7 and single-dose vaccinated (green) had higher nAb IC_{50} against B.1.1.7 compared with two-dose vaccinated infection-naïve HCWs ($P < 0.0001$) (blue) and Wuhan Hu-1 previously infected single-dose vaccinated HCWs ($P < 0.0001$) (red) (Fig. 3G and fig. S10A). Furthermore, neutralization IC_{50} of B.1.351 live virus was less potent after heterologous B.1.1.7 infection and single-dose vaccination (green) ($P = 0.0001$; Fig. 3G and fig. S10A). After three antigen encounters, the impact of heterologous exposure became more pronounced. HCWs with heterologous infection with B.1.1.7 (green) rather than Wuhan Hu-1 (red) had substantially lower neutralization against authentic Wuhan Hu-1 ($P < 0.0001$), B.1.351 ($P < 0.0001$), and P.1 ($P = 0.0616$) live virus (fig. S10B). Furthermore,
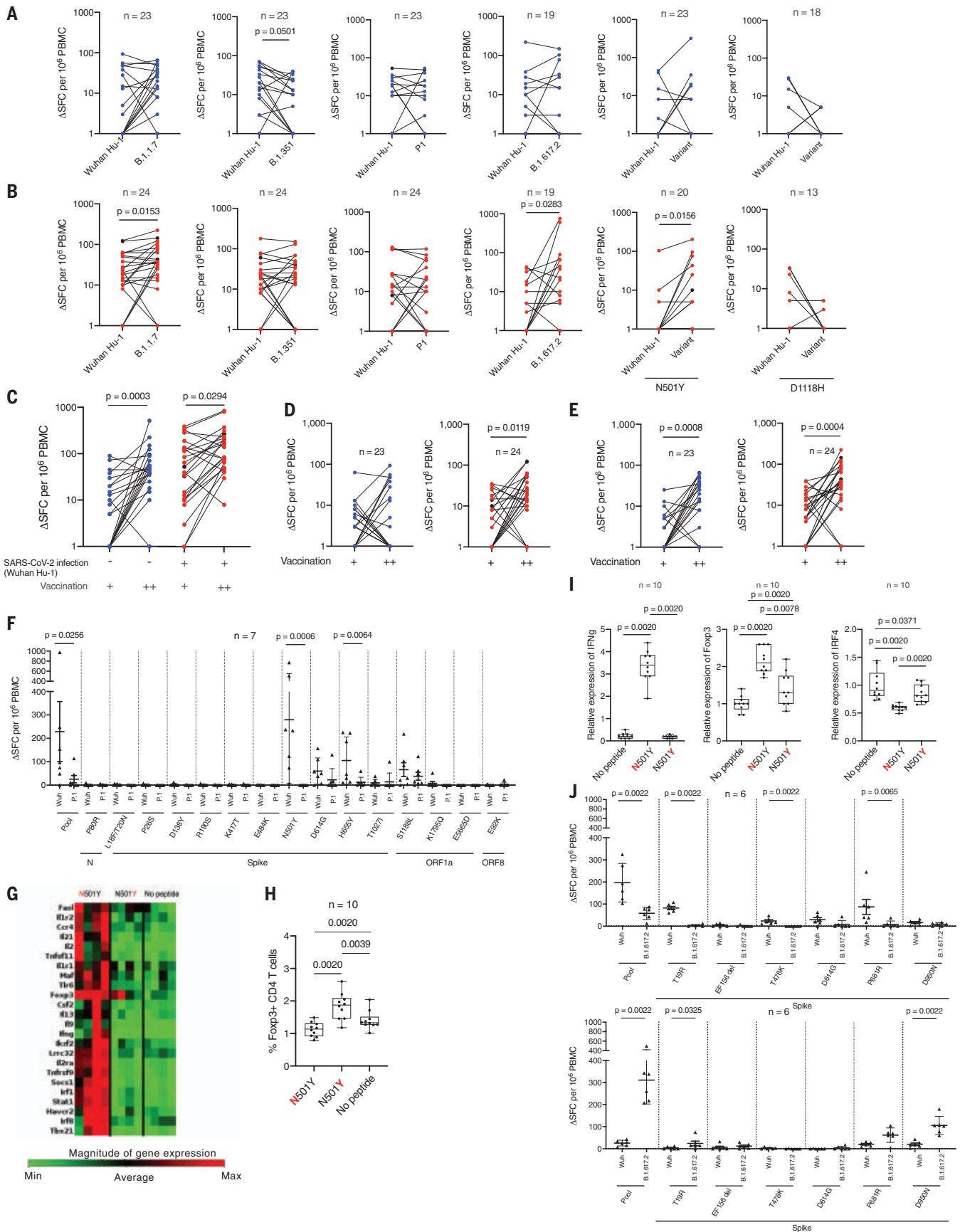
Fig. 2. T cell responses to Wuhan Hu-1 and VOC peptide pools after one- and two-dose BNT162b2 vaccination in infection-naïve and previously infected HCWs. Magnitude of T cell responses to B.1.1.7, B.1.351, P.1, and B.1.617.2 variant specific peptide pools and to the matched sequence peptide pools from Wuhan Hu-1 as well as to individual wild-type and variant N501Y and D1118H peptides in (**A**) SARS-Cov-2 infection-naïve (blue) HCWs and (**B**) previously infected (red) HCWs who had received two doses of BNT162b2 vaccine. In all figures, HCWs without labaoratory-confirmed infection are shown in blue, HCWs with laboratory-confirmed infection are shown in red, and HCWs with new infection or reinfection after 30 weeks are shown in black. (**C** to **E**) Magnitude of T cell responses to spike MEP (C), Wuhan Hu-1 (D), and B.1.1.7 variant (E) peptide pools after one (+) and two (++) doses of BNT162b2 vaccine in infection-naïve HCWs (blue) and previously infected HCWs (red). (**F**) Magnitude of T cell response to wild-type Wuhan Hu-1 and P.1 variant peptide pools and individual peptides in Wuhan Hu-1 peptide pool immunized HLA-DRB1*04:01 transgenic mice (n = 7). (**G**) Heatmap showing relative gene expression of T cell activation markers in draining lymph node (DLN) cells from Wuhan Hu-1 N501Y peptide–primed DRB1*04:01 transgenic mice (n = 4) stimulated for 24 h in vitro with 10 µg/ml of wild-type Wuhan Hu–1 or variant peptide. Genes shown in red are significantly different (P < 0.05) between no peptide control and Wuhan Hu-1 or variant peptide stimulated cells with a fold change greater than 1.5. (**H**) Percentage of Foxp3+ CD4 T cells by flow cytometry in DLN cells from Wuhan Hu-1 N501Y peptide-primed DRB1*04:01 transgenic mice (n = 10) stimulated for 24 h in vitro with 10 mg/ml of wild-type or variant peptide. (**I**) Relative gene expression of *ifng*, *foxp3*, and *irf4* in DLN cells from Wuhan Hu-1 N501Y peptide-primed DRB1*0401 transgenic mice (n = 10) stimulated for 24 h in vitro with 10 µg/ml of Wuhan Hu-1 or variant peptide. (**J**) Magnitude of T cell response to wild-type Wuhan Hu-1 and B.1.617.2 variant peptide pools and individual peptides in Wuhan Hu-1 peptide pool immunized (top panel, n = 6) or B.1.617.2 peptide pool immunized (bottom panel, n = 6) HLA-DRB1*04:01 transgenic mice. In (A) to (E), (I), and (J), Wilcoxon matched-pairs signed rank test was used. New infection (in the infection-naïve group, n = 2) and reinfection (in the previously infected group, n = 2) data points shown in black were excluded from statistical analysis. In (F) and (G), Mann-Whitney U test was used.

two-dose vaccinated HCWs with a history of heterologous infection by B.1.1.7 had lower nAb IC_{50}s against B.1.351 than infection-naïve double vaccinated HCWs (P = 0.0007; fig. S10B). By contrast, two-dose vaccinated HCWs with a history of heterologous infection with B.1.1.7 (green) had higher nAb IC_{50} against B.1.617.2 than infection-naïve two-dose vaccinated HCWs (P < 0.0001; fig. S10B). Thus, the neutralization potency achieved was dependent on the number of antigen encounters and the presence or absence of heterologous exposure. Ab neutralization is a widely accepted correlate of protection (COP), estimated at an nAb IC_{50} of >15, and S1 RBD binding is an accessible and accepted marker of this (22). Our data here show that S1 RBD and VOC S1 RBD titers are not a reliable COP for VOC, especially in the context of selective discriminative heterologous exposure during infection with B.1.1.7. In HCWs with heterologous B.1.1.7 exposure, the nAb IC_{50} of B1.351 was significantly reduced (P < 0.0001) and B.1.617.2 was significantly increased (P = 0.0025) compared with HCWs with homologous Wuhan Hu-1 infection.

Differential longitudinal immunity after heterologous infection

Next, we considered the relative durability of immunity after three-dose homologous (Wuhan Hu-1 infection plus vaccination) and heterologous (B.1.1.7 infection plus vaccination) antigen compared with two-dose vaccinated infection-naïve individuals. S1 RBD Ab levels declined as the number of weeks after the second dose increased in both infection-naïve (r = 0.545; P < 0.0001) and previously infected (r = 0.537; P < 0.0001) HCWs (Fig. 4, A and B, and tables S1, S5 to S7, and S9). After a third vaccine dose, S1 RBD binding increased in all three groups to a similar level (Fig. 4C). B.1.1.7 infection (green circles) compared with Wuhan Hu-1 infection (red circles) was associated with lower nAb IC_{50} against Wuhan Hu-1 (P < 0.0001) and

B.1.351 (P < 0.0001) VOC initially at (10 d to 8 weeks after vaccination) and P.1 (P = 0.0020) 8 weeks later (12 to 22 weeks after vaccination), indicating that differential nAb responses over time depend on the VOC (Fig. 4D). Two-dose vaccinated infection-naïve HCWs were at a lower starting value for nAb IC_{50} against B.1.617.2 compared with HCWs infected with B.1.1.7 (P < 0.0001) and Wuhan Hu-1 (P = 0.0462) (Fig. 3H and fig. S10B). The nAb IC_{50} against authentic Wuhan Hu-1 and all four VOCs (B.1.1.7, B.1.351, P.1, and B.1.617.2) fell differentially depending on the VOC being neutralized and SARS-CoV-2 infection history (Fig. 4E).

In two-dose vaccinated infection-naïve HCW, nAb IC_{50} against authentic Wuhan Hu-1 and all four of the VOCs fell significantly over 18 weeks (Wuhan Hu-1, 26-fold, P < 0.0001; B.1.1.7, 18-fold, P < 0.0001; B.1.351, 29-fold, P < 0.0001; P.1, 26-fold, P < 0.0001; B.1.617.2 233-fold, P < 0.0001). This was most marked for B.1.617.2 live virus neutralization; by 21 weeks after the second dose, it was zero for 21/27 (78%) HCWs tested (median = 0, IQR = 0, n = 27) (Fig. 4E). For two-dose vaccinated HCWs previously infected with Wuhan Hu-1, there was a significantly and less pronounced fall in nAb IC_{50} against authentic Wuhan Hu-1 and three of the VOCs over 18 weeks (Wuhan Hu-1, fivefold, P = 0.0003; B.1.1.7, sevenfold, P < 0.0001; B.1.351, fourfold, P = 0.0042; P.1, 10-fold, P = 0.0003). By contrast, in the presence of heterologous infection with B.1.1.7 and two-dose vaccination, nAb IC_{50} fell significantly over 18 weeks for some variants (Wuhan Hu-1, sixfold, P = 0.0002; B.1.1.7 13-fold, P < 0.0001; P.1 sevenfold, P < 0.0001; and B.1.617.2 41-fold, P < 0.0001), whereas nAb IC_{50} against B.1.351 did not (P = 0.4304).

T cell responses against spike MEP pool 21 weeks after the second dose were higher in two-dose vaccinated HCWs that were infected with B.1.1.7 than in two-dose vaccinated infection-naïve HCWs (P = 0.0304; Fig. 4F). MBC frequencies against the Wuhan Hu-1 S1 and S1 containing the B.1.617.2 mutations

were equivalent 21 weeks after second dose vaccination (Fig. 4G). Two-dose vaccinated B.1.1.7 infected HCW had higher frequency responses against S1 containing the B.1.617.2 mutations than two-dose vaccinated infection-naïve HCWs (P = 0.0138; Fig. 4H). MBC frequencies in two-dose vaccinated infection-naïve HCWs fell over 18 weeks (n = 11, P = 0.0098), whereas those from Wuhan Hu-1 previously infected HCWs were sustained (n = 4; Fig. 4I).

Immune parameters associated with B.1.617.2 breakthrough infection

At the time of the 71- to 72- and 83- to 84-week recruitments, B.1.617.2 VOC infections accounted for 97.8 and 99.9% of UK SARS-CoV-2 infections, respectively (26). We identified 6/80 (8%) and 14/74 (19%) B.1.617.2 breakthrough infections in the double-vaccinated HCWs. Ten were previously infection-naïve, four had previous Wuhan Hu-1 infection, and six were infected with B.1.1.7 (tables S8 and S10). PCR-positive confirmed B.1.617.2 breakthrough infections occurred in the context of S1 RBD Ab levels ranging 1110 to 29,308 U/ml (median = 9010, IQR = 13,650) 2 to 3 weeks after the second vaccine dose. We evaluated the S1 RBD, nAb, T cell, and MBC data in these double-vaccinated HCW before (purple circle) and after (lilac circle) their B.1.617.2 breakthrough infection (fig. S11, A to J). Analysis showed that both previously infected and infection-naïve HCWs with S1 RBD Ab responses (20 d after the second dose) of >1100 U/ml had become infected, and breakthrough infection was not specifically associated with being a low responder (fig. S11, A and B). Relatively potent nAb IC_{50} at 20 d after the second vaccine dose had fallen 12 weeks later. This fall was more pronounced in the infection-naïve double-vaccinated HCWs (fig. S11, D and E). For B.1.617.2, the nAb IC_{50} fell to zero in two-dose vaccinated infection-naïve HCWs (fig. S11, D and G). B.1.617.2 breakthrough infections have been linked to low nAb responses (33). Current
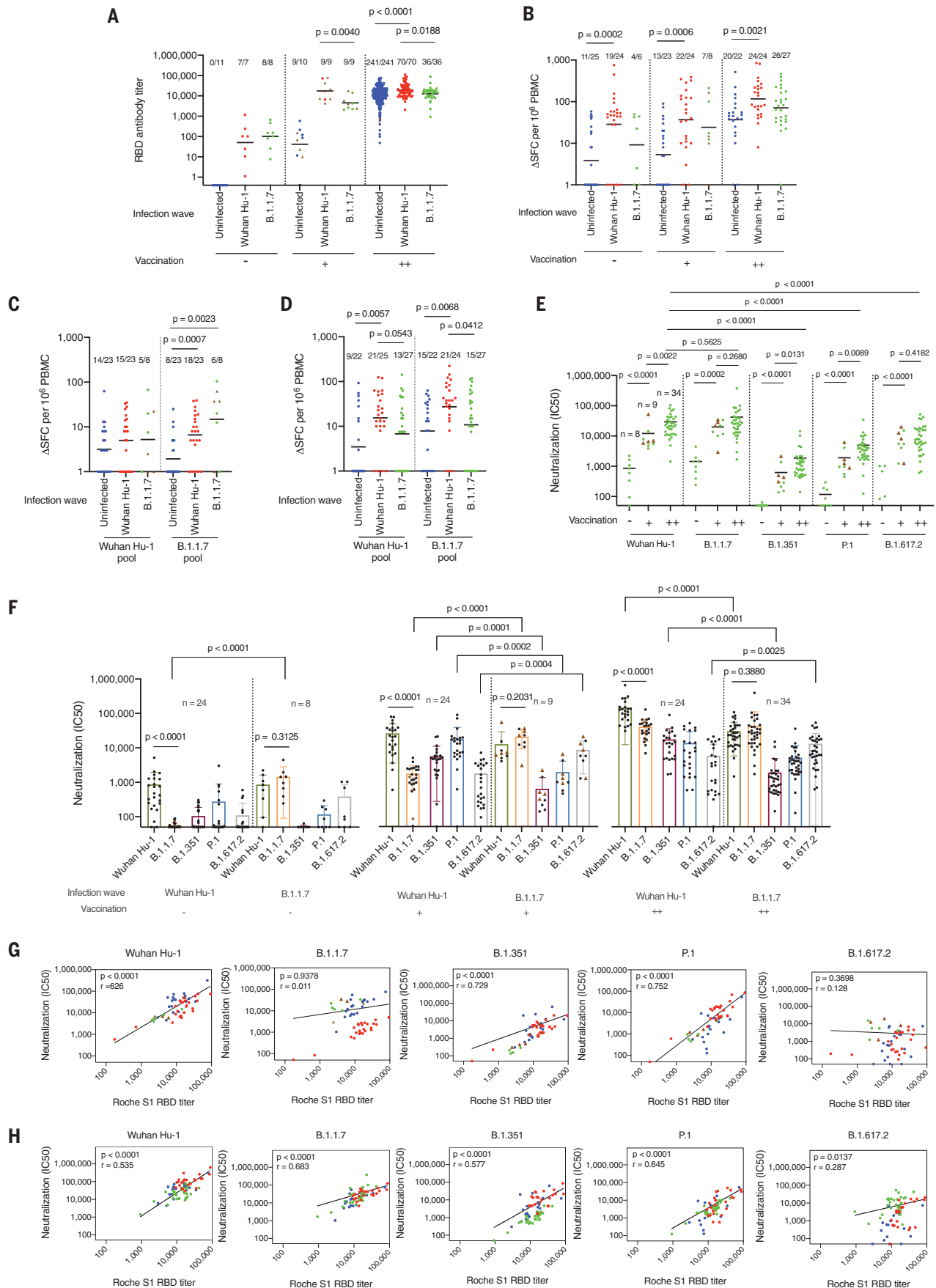
Fig. 3. T cell and B cell immunity and neutralization hierarchy after heterologous exposure through infection with the B.1.1.7 VOC during the second UK wave and in the context of single- and two-dose vaccination. (**A**) RBD Ab titers at 54 to 57 weeks after the start of study recruitment in March 2020 in HCWs who were not infected with SARS-CoV-2 (blue, $n = 256$), those infected during the first UK wave by the Wuhan Hu-1 strain (red, $n = 86$), and those infected during the second UK wave by the B.1.1.7 (green, $n = 53$). (**B**) Magnitude of T cell response to Wuhan Hu-1 spike MEP peptide pool. Data are plotted according to whether individuals were unvaccinated (−) or had received one (+) or two (++) doses of BNT162b2 vaccine. (**C** and **D**) Magnitude of T cell response to B.1.1.7 peptide pools (Wuhan Hu-1 or B.1.1.7 variant peptides) after first dose (C) or second dose (D) of vaccine. (**E** and **F**) Neutralizing Ab titers (IC_{50}) against authentic Wuhan Hu-1 live virus and B.1.1.7, B.1.351, P.1, or B.1.617.2 VOCs in HCWs infected with SARS-CoV-2 during

the B.1.1.7 wave (E) and during the Wuhan Hu-1 first UK wave (F), plotted according to whether individuals were unvaccinated (−, Wuhan Hu-1 wave, $n = 24$; B.1.1.7 wave, $n = 8$) or had received one dose (+, Wuhan Hu-1 wave $n = 24$, B.1.1.7 wave $n = 9$) or two doses (++, Wuhan Hu-1 wave $n = 24$, B.1.1.7 wave $n = 34$) of vaccine. (**G** and **H**) Correlation between Roche S1 RBD Ab titer and nAb (IC_{50}) against authentic Wuhan Hu-1 live virus and B.1.1.7, B.1.351, P.1, and B.1.617.2 VOC in one-dose (G) and two-dose (H) BNT162b2 vaccinated HCWs previously infected by Wuhan Hu-1 (red, $n = 23$) or B.1.1.7 infected (green, one dose $n = 9$, two doses $n = 31$) and two-dose vaccinated infection-naïve (blue, $n = 19$) HCWs at 54 to 57 weeks after initial study recruitment. In all graphs, individuals who received the ChAdOx1 nCoV-19 vaccine are marked as brown triangles. In (A) to (E), Mann-Whitney U test was used; in (F), Mann-Whitney U test or Wilcoxon matched-pairs signed rank test was used; and in (G) and (H), Spearman's rank correlation was used.

estimates for COP apply poorly in the case of B.1.617.2 (22). In the cases reported here, breakthrough infections occurred in the face of good prior S1 RBD Ab titers, nAb IC_{50}s, and MBC frequency 20 d after second vaccine dose, but responses fell in the subsequent 18 weeks, especially in the infection-naïve two-dose vaccinated HCWs. Initial immune responses poorly predicted protection against B.1.617.2 breakthrough infection (fig. S11, A to J).

Discussion

The new challenge in a world facing diverse VOCs is to understand both cross-neutralization of these by first-generation spike vaccines and also how protective immunity is differentially shaped in those who have had infection by the different VOCs (34). These real-life issues of immune imprinting have implications for the optimal design of second-generation variant spike vaccine boosters in parts of the world experiencing different VOCs or for assessing heterogeneity in the quantity, quality, and durability of immune protection in those who have experienced different combinations of infecting and vaccinating sequences. Infection by B.1.1.7 has different features from Wuhan Hu-1 in terms of protection against VOCs, which is evident in the response to infection and in the differential impact of vaccine-boosting previously described (8, 9, 11–14). Differential skewing of the immune repertoire may occur if the prime and/or boost is through B.1.1.7, because epitope modification at DY144 and N501Y changes the nAb repertoire (4, 35). We found a differential hierarchy of cross-neutralization to VOCs after natural infection or one or two vaccine doses, depending on the infecting strain. That different immune priming by B.1.1.7 is likely a consequence of heterologous virus per se and not of a shorter time span for affinity maturation is supported by data from Israel, confirming that susceptibility grows with time and with declining Ab levels (36). Vaccine programs currently use a prototypic Wuhan Hu-1 sequence against a background of infection by different VOCs predominating in different parts of the world.

Decoding the differential breadth of VOC-neutralizing responses ensuing from diverse priming combinations will affect which variant spike sequences may best serve in second-generation vaccines. This appears to be a more complex choice than simply opting for the most concerning variant at any given time. Because heterologous combinations can confer a diminished response against other variants, there may be a case for sticking with the Wuhan Hu-1 sequence in booster vaccinations in the first instance in the face of unknown future VOCs or for improved efforts to define vaccines based on optimization of common, conserved neutralizing epitopes (37). In any case, the inference from this cohort is that populations infected during waves of different variants carry distinct immune memory, with implications for differential protection against future VOCs.

Methods summary

Detailed materials and methods are provided in the supplementary materials. Recruitment of 731 HCWs into the COVIDsortium cohort followed longitudinally with weekly self-reported symptom diaries, SARS-CoV-2 PCR, N, and S1 RBD serology for 16 weeks from the start of the first UK wave has been described previously (5–10). This enabled a cross-sectional, case-controlled sub-study of 136 HCWs recruited 16 to 18 weeks after the March 2020 UK lockdown that reported discordant neutralizing antibody and T cell responses in SARS-CoV-2 natural infection during the first wave (Wuhan Hu-1) (6). A cross-sectional, case-controlled vaccine sub-study cohort of 51 HCWs at 22 d after first BNT162b2 dose described vaccine immunity in HCWs with ($n = 25$) and without ($n = 26$) prior SARS-CoV-2 infection during the Wuhan Hu-1 wave (8, 9). The current sub-study includes longitudinal follow-up of this previously published vaccine sub-study cohort ($n = 51$) at 20 d (IQR = 7) after a second BNT162b2 dose plus an additional 358 HCWs, 53 of whom were infected with B.1.1.7 during the second UK wave. At 71 to 72 weeks, 80 two-dose vaccinated HCWs were re-recruited 18 to 21 weeks after their second dose for longitudi-

nal follow-up of SARS-CoV-2 infection-naïve ($n = 27$) or HCWs previously infected during the Wuhan Hu-1 UK wave ($n = 31$) or the second B.1.1.7 UK wave ($n = 22$). At 83 to 84 weeks, 74 previously two-dose vaccinated HCWs who were infection naïve ($n = 30$) or infected during the Wuhan Hu-1 wave ($n = 18$) or the B.1.1.7 wave ($n = 19$) were re-recruited 30 to 33 weeks after their second dose. Sixty-seven (91%) were within a median of 18 d (IQR = 12) after their third BNT162b2 dose. HCWs with breakthrough infection by B.1.617.2 were identified by PCR and N serology. Peripheral blood mononuclear cells and serum samples were prepared and cryopreserved as previously described (6, 9). Anti-nucleocapsid [cutoff index (COI) ≥ 1.0 U/ml, positive] and anti-spike antibody (COI ≥ 0.8 U/ml, positive) detection Ab testing was conducted at UK Health Security Agency (UKHSA) using the Roche Cobas e801 analyzer. Recombinant proteins were used in ELISAs looking at the VOC S1 RBD responses: SARS-CoV-2 spike glycoprotein (S1) RBD, SARS-CoV-2 (N501Y mutant), (501Y.V2: K417N, E484K, N501Y), (B.1.1.28: K417T, E484K, N501Y), and (B.1.617.2: L452R, T478K) spike glycoprotein (S1) RBDs derived from Wuhan Hu-1, B.1.1.7, B.1.351, P.1, and B.1.617.2 VOC, respectively. T cell experiments peptide panels included an MEP composed of a pool of 18 12- to 20-mer peptide epitopes (6, 9), and VOC pools contained peptides from the B.1.1.7, B.1.351, P.1, and B.1.617.2 sequences and their respective Wuhan Hu-1 sequence (table S2). IFNγ-T cell ELISpots and MBC ELISpots were performed as previously described (6, 9). MBC ELISpot plates were coated with phosphate-buffered saline, purified anti-human IgG MT91/145, SARS-CoV-2 S1 spike, E484K, K417N, N501Y spike or T19R, G142D, del 156-157, R158G, L452R, T478K, D614G, P681R spike. In silico predictions of HLA-DRB1/peptide binding were performed using NetMHCIIpan-4.0 (9, 38, 39). Studies using HLAII transgenics (DRB1*0401) were performed as previously described (9, 40, 41). For transcriptomic and flow cytometry analysis, mouse lymph node cells were cultured with no peptide and wild-type or variant N501Y
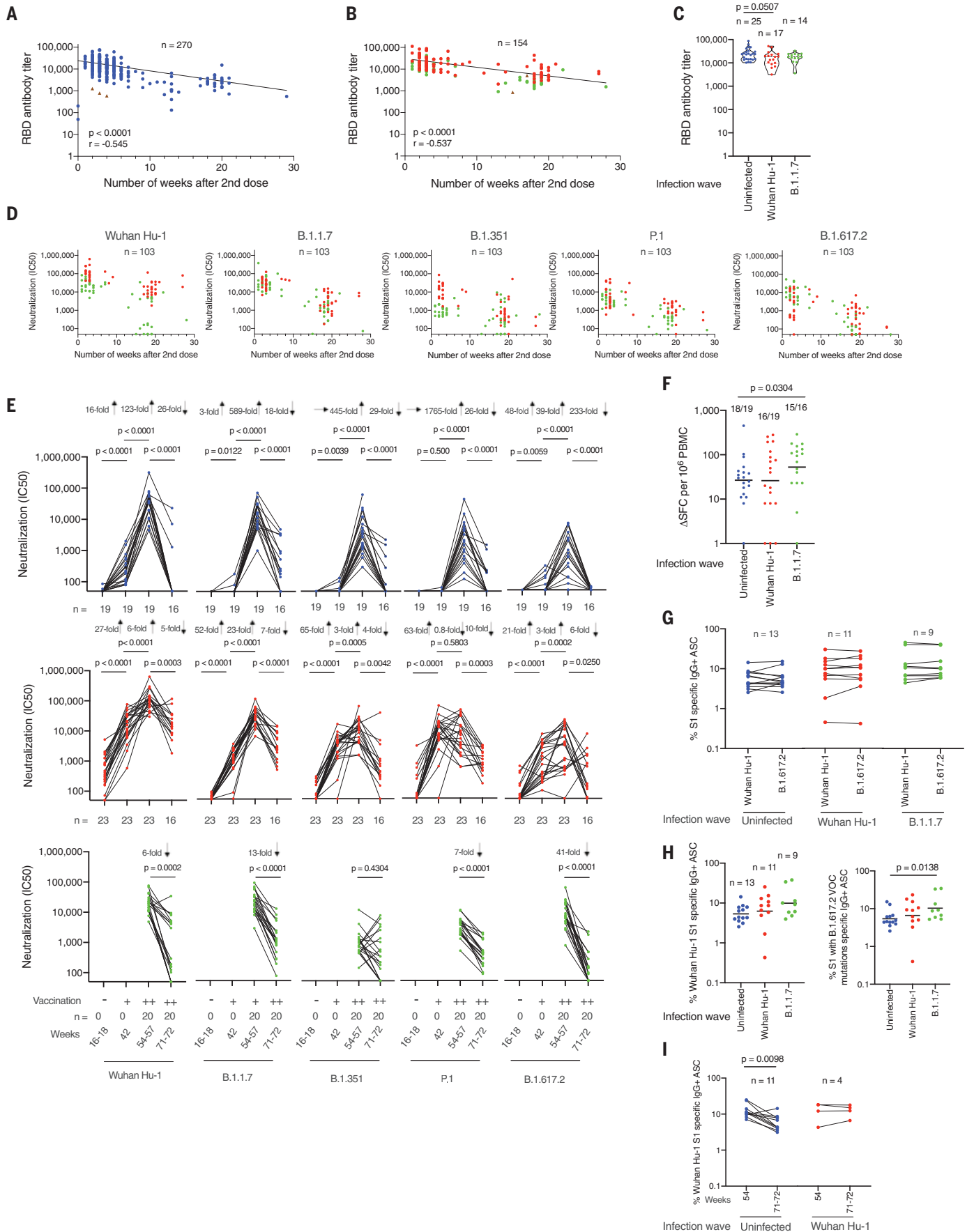
Fig. 4. Differential impact of heterologous exposure through B.1.1.7 infection on neutralization of other VOC and durability of the immune response. (**A**) S1 RBD Ab titers in HCWs with no previous SARS-CoV-2 infection (blue) and two-dose vaccination (n = 270). (**B**) S1 RBD Ab titers in HCWs with a history of SARS-CoV-2 infection (Wuhan Hu-1, red; B.1.17, green) and two-dose vaccination (n = 154), plotted by the number of weeks that serum was sampled after the second vaccine dose. HCWs who received the ChAdOx1 nCoV-19 vaccine are indicated by brown triangles. (**C**) S1 RBD Ab titers in infection-naïve HCWs (blue, n = 25) or those with a history of SARS-CoV-2 infection with the Wuhan Hu-1 strain (red, n = 17) or B.1.17 VOCs (green, n = 14) and three-dose vaccination sampled between 10 d and 7 weeks after the third vaccine dose. (**D**) Neutralizing Ab titers (IC_{50}) against authentic Wuhan Hu-1 live virus and B.1.1.7, B.1.351, P.1, and B.1.617.2 VOC in HCWs with a history of previous SARS-CoV-2 infection who had received two vaccine doses, plotted by the number of weeks that serum was sampled after the second vaccine dose. (**E**) Neutralizing Ab titer (IC_{50}) against authentic Wuhan Hu-1 live virus and B.1.1.7, B.1.351, P.1, and B.1.617.2 VOCs plotted longitudinally 20 d after the first (+) and 20 days after the second (++) dose and then 21 weeks after the second BNT162b2 vaccination. Shown are SARS-CoV-2 infection-naïve HCWs (blue, upper panel), Wuhan Hu-1 previously infected HCWs (red, middle panel),

and B.1.1.7–infected HCWs (green, lower panel). (**F**) Magnitude of T cell response to spike MEP peptide pool at 71 to 72 weeks after initial study recruitment in double-vaccinated infection-naïve HCWs (blue, n = 19) and HCWs with a history of SARS-CoV-2 infection by Wuhan Hu-1 (red, n = 19) or B.1.1.7 (green, n = 16). (**G**) Percentage of IgG^+ ASCs specific for Wuhan Hu-1 S1 or S1 protein containing T19R, G142D, del 156-157, R158G, L452R, T478K, D614G, and P681R (B.1.617.2 VOC) mutations at 71 to 72 weeks after initial study recruitment in double-vaccinated infection-naïve (blue, n = 13) HCWs and those with a history of SARS-CoV-2 infection with the Wuhan Hu-1 strain (red, n = 11) and B.1.1.7 VOC (green, n = 9) 21 weeks after the second vaccine dose. (**H**) Percentage of IgG^+ ASCs specific for Wuhan Hu-1 S1 protein (left panel) or S1 protein containing T19R, G142D, del 156-157, R158G, L452R, T478K, D614G, and P681R (B.1.617.2 VOC) mutations (right panel) at 71 to 72 weeks after initial study recruitment in double-vaccinated infection-naïve HCWs (blue, n = 13), previously Wuhan Hu-1–infected HCWs (red, n = 11), and B.1.1.7–infected HCWs (green, n = 9). (**I**) Percentage of IgG^+ ASCs specific for Wuhan Hu-1 S1 protein in double-vaccinated HCWs with (red, n = 4) or without (blue, n = 11) a history of Wuhan Hu-1 infection at 54 and 71 to 72 weeks after initial study recruitment. In (A) and (B), Spearman's rank correlation was used; in (C), (F), and (H), Mann-Whitney U test was used; and in (E) and (I), Wilcoxon matched-pairs signed rank test was used.

peptides, and at 24 hours, cells were harvested and lysed for RNA extraction or stained for flow cytometry. SARS-CoV-2 microneutralization assays were performed as described previously using VeroE6 cells (6, 9). Participant sera were incubated with 3×10^4 focus-forming units of SARS-CoV-2 (Wuhan Hu-1, B.1.1.7, B.1.351, P.1, or B.1617.2 isolates of SARS-CoV-2 live virus).

REFERENCES AND NOTES

1. D. A. Collier et al., *Nature* **596**, 417–422 (2021).
2. E. C. Wall et al., *Lancet* **397**, 2331–2333 (2021).
3. D. Planas et al., *Nature* **596**, 276–280 (2021).
4. W. T. Harvey et al., *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
5. T. A. Treibel et al., *Lancet* **395**, 1608–1610 (2021).
6. C. J. Reynolds et al., *Sci. Immunol.* **5**, eabf3698 (2020).
7. C. Manisty et al., *EBioMedicine* **65**, 103259 (2021).
8. C. Manisty et al., *Lancet* **397**, 1057–1058 (2021).
9. C. J. Reynolds et al., *Science* **372**, eabh1282 (2021).
10. R. K. Gupta et al., *Lancet Microbe* **2**, e508–e517 (2021).
11. F. Krammer et al., *N. Engl. J. Med.* **384**, 1372–1374 (2021).
12. L. Stamatatos et al., *Science* **372**, eabg9175 (2021).
13. R. R. Goel et al., *Sci. Immunol.* **6**, eabi6950 (2021).
14. J. E. Ebinger et al., *Nat. Med.* **27**, 981–984 (2021).
15. P. S. Arunachalam et al., *Nature* **596**, 410–416 (2021).
16. A. Zhang, H. D. Stacey, C. E. Mullarkey, M. S. Miller, *J. Immunol.* **202**, 335–340 (2019).
17. T. Aydillo et al., *Nat. Commun.* **12**, 3781 (2021).
18. A. K. Wheatley et al., *Trends Immunol.* **42**, 956–959 (2021).
19. N. Kamar et al., *N. Engl. J. Med.* **385**, 661–662 (2021).
20. A. Del Bello et al., *Am. J. Transplant.* ajt.16775 (2021).
21. D. M. Altmann, R. J. Boyton, R. Beale, *Science* **371**, 1103–1104 (2021).
22. S. Khoury et al., *Nat. Med.* **27**, 1205–1211 (2021).
23. A. M. Solinger, M. E. Ultee, E. Margoliash, R. H. Schwartz, *J. Exp. Med.* **150**, 830–848 (1979).
24. D. P. Martin et al., *Cell* **184**, 5189–5200.e7 (2021).
25. L. Lu et al., *EBioMedicine* **71**, 103544 (2021).
26. GISAID, "Tracking of variants" (2021); https://www.gisaid.org/hcov19-variants/
27. T. M. Snyder et al., Magnitude and dynamics of the T-cell response to SARS-CoV-2 Infection at both individual and population levels. medRxiv 2020.07.31.20165647 [Preprint] (2020); https://doi.org/10.1101/2020.07.31.20165647.
28. Wellcome Sanger Institute, "COVID-19 genomic surveillance" (2021); https://covid19.sanger.ac.uk/lineages/raw.
29. F. Muecksch et al., *Immunity* **54**, 1853–1868.e7 (2021).
30. Z. Wang et al., *Nature* **595**, 426–431 (2021).
31. C. Gaebler et al., *Nature* **591**, 639–644 (2021).
32. N. Faulkner et al., *eLife* **10**, e69317 (2021).
33. M. Bergwerk et al., *N. Engl. J. Med.* **385**, 1474–1484 (2021).
34. T. Moyo-Gwete et al., *N. Engl. J. Med.* **384**, 2161–2163 (2021).
35. M. McCallum et al., *Cell* **184**, 2332–2347.e16 (2021).
36. Y. Goldberg et al., *N. Engl. J. Med.* NEJMoa2114228 (2021).
37. L. A. VanBlargan et al., *Immunity* **54**, 2399–2416.e6 (2021).
38. D. M. Altmann, C. J. Reynolds, R. J. Boyton, *Cell Rep. Med.* **2**, 100286 (2021).
39. B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, *Nucleic Acids Res.* **48** (W1), W449–W454 (2020).
40. K. J. Quigley et al., *Am. J. Respir. Crit. Care Med.* **191**, 1250–1264 (2015).
41. C. Reynolds et al., *J. Immunol.* **194**, 4814–4824 (2015).
42. K. Tao et al., *Nat. Rev. Genet.* **22**, 757–773 (2021).
43. J. Zou et al., *NPJ Vaccines* **6**, 44 (2021).

REPORTS

TOPOLOGICAL MATTER

Distinguishing between non-abelian topological orders in a quantum Hall system

Bivas Dutta, Wenmin Yang, Ron Melcer, Hemanta Kumar Kundu, Moty Heiblum*, Vladimir Umansky, Yuval Oreg, Ady Stern, David Mross

Quantum Hall states can harbor exotic quantum phases. The nature of these states is reflected in the gapless edge modes owing to "bulk-edge" correspondence. The most studied putative non-abelian state is the spin-polarized filling factor (ν) = 5/2, which permits different topological orders that can be abelian or non-abelian. We developed a method that interfaces the studied quantum state with another state and used it to identify the topological order of ν = 5/2 state. The interface between two half-planes, one hosting the ν = 5/2 state and the other an integer ν = 3 state, supports a fractional ν = 1/2 charge mode and a neutral Majorana mode. The counterpropagating chirality of the Majorana mode, probed by measuring partition noise, is consistent with the particle-hole Pfaffian (PH-Pf) topological order and rules out the anti-Pfaffian order.

The quantum Hall effect (QHE) state harbors an insulating bulk and conductive edges and is the earliest known example of a topological insulator (1). It is characterized by topological invariants, which are stable to small changes in the details of the system (2). Of these quantities, the easiest to probe is the electrical Hall conductance of the edge mode, $G_H = \nu e^2/h$, where ν is the bulk filling factor (integer or fractional), e is the electron charge, and h is Planck's constant. Quantum Hall (QH) states with fractional filling factors support quasiparticles with fractional charges and anyonic statistics. The ubiquitous Laughlin states are abelian; exchanging the positions of their quasiparticles adds a phase to the ground-state wave function (3–5). In the more exotic non-abelian states (6, 7), the presence of certain quasiparticles results in multiple degenerate ground states, and interchanging these quasiparticles cycles the system between the different ground states. In general, QH states permit different topological orders, and the usual conductance and charge measurements are not sufficient to distinguish between the different topological orders.

Another topological invariant, the thermal QH conductance G_T, may help make that distinction. The thermal conductance is sensitive to all energy-carrying edge modes, charged or neutral, and can be expressed as $G_T = KT$, where K is the thermal conductance coefficient and T is the temperature. It was proposed (7, 8) and experimentally proven that K of a single chiral and ballistic mode—fermionic

Braun Center for Sub-Micron Research, Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel.
*Corresponding author. Email: moty.heiblum@weizmann.ac.il

(9), bosonic (10, 11), or (abelian) anyonic (12)—is quantized $K = \kappa_0$, with $\kappa_0 = \pi^2 k_B^2/3h$, where k_B is the Boltzmann constant. However, a fractional value of K is expected for non-abelian states (13). For the ν = 5/2 state, we found a thermal Hall conductance coefficient, $K = 2.5\kappa_0$, that is consistent with the non-abelian particle-hole Pfaffian (PH-Pf) topological order (14).

However, there is a caveat: For a QH state that supports multiple edge modes, some moving downstream (DS; in the chirality dictated by the magnetic field) and some upstream (US; in opposite chirality), the theoretically predicted thermal conductance assumes a full thermal equilibration among all modes (15–17). For example, for modes of integer thermal conductance, the predicted value is $G_T = (n_d - n_u)\kappa_0 T$, where n_d and n_u are the number of DS and US modes, respectively. In the other extreme, with no thermal equilibration, $G_T = (n_d + n_u)\kappa_0 T$ (12). Therefore, the thermal equilibration length, which is in general longer than the charge equilibration length, is of crucial importance in interpreting thermal conductance measurements.

Moore and Read predicted that the topological order of the ν = 5/2 state is a Pfaffian (Pf) state, with $K = 3.5\kappa_0$ (18). However, the Pf mode was ruled out when a US neutral mode was observed experimentally (19) because the Pf order does not support a topologically protected US mode. The "particle-hole conjugate" of the Pf order, the anti-Pfaffian (A-Pf) with $K = 1.5\kappa_0$ (20, 21), does exhibit a US mode. Numerical studies found both Pf and A-Pf to be highly competitive ground states in a homogenous system (neglecting disorder). The two are degenerate within a single Landau level, and Landau level mixing may tip the balance between the two in either direction

(22, 23). Several theoretical proposals offer possible explanations for the discrepancy between numerical calculations and the experimentally found PH-Pf order: (i) Inhomogeneity in the density may lead to islands of local Pf and A-Pf orders, from which a global PH-Pf order emerges (24–28). (ii) A considerably longer thermal equilibration length than the size of the device may lead to deviation from the expected theoretical value. In particular, an unequilibrated Majorana mode in the A-Pf (15, 29) order will add its contribution to K instead of subtracting it, leading to $K = 2.5\kappa_0$ (30–32).

One may suggest to measure the thermal conductance at a short propagation length, where thermal equilibration is practically negligible, expecting $K = (n_d + n_u)\kappa_0$; namely, $K_{A-Pf} = 4.5\kappa_0$ or $K_{PH-Pf} = 3.5\kappa_0$. However, "spontaneous edge reconstruction" may add short-lived pairs of counterpropagating (non-topological) neutral modes (33), thus increasing the apparent thermal conductance of the states at short distances.

We developed a method in which a studied state is interfaced with another state, with the interface harboring an isolated edge channel. We used this method to probe the Majorana modes' chirality at interfaces between the ν = 5/2 state and ν = 2 and ν = 3 integer states. Measurements fulfill two important requirements: (i) thermally unequilibrated transport, in which measurements are performed in the intermediate length regime—longer than the charge equilibration length (leading to the quantized interface conductance) but shorter than the thermal equilibration length (allowing unequilibrated modes transport) (15, 29, 34)—and (ii) compensating integer modes, in which the counterpropagating integer modes at the interface between ν = 5/2 and ν = 2 or 3 mutually localize each other. This eliminates spurious hotspots and allows the identification of the topological order.

We used gated high-quality GaAs-AlGaAs molecular beam epitaxy (MBE)–grown heterostructures (35). Structures were designed to resolve the contradictory requirements of the doped layers, which should ensure a full quantization of the fragile ν = 5/2 state and, at the same time, allow highly stable and "hysteresis-free" operation of the gated structures (14). Two separate gates divided the surface into two, an upper half and lower half (Fig. 1A). The gates were isolated from the sample and from each other by 15- to 25-nm layers of HfO$_2$ [more details of the structure are provided in (35)]. A gate-voltage in the range of –1.5 V < V_g < +0.3 V allows varying the electron density from pinch off to 3×10^{11} cm^{-2} (fig. S1). Tuning each gate separately controls the two interfaced filling factors, leading to the desired interface modes' conductance. The ohmic contacts at the physical edge of the mesa probe the filling

factors of the two respective half-planes, and the contacts at the interface probe the interface.

The heart of the measurement setup (Fig. 1A, dashed box) consists of three ohmic contacts at the interface, with the source (Fig. 1A, "S"), placed symmetrically between the amplifier contact (Fig. 1A, "A") and the cold-grounded drain (Fig. 1A, "D"). An injected dc source current at the interface, I_S, forms a hotspot at the back side of the source (and at the front side of the drain) (36). In states that support counterpropagating modes, the thermally activated modes by the hotspot (usually, neutral modes), lead to shot noise at the am-

plifier contact (37, 38). The noise was filtered with an LC resonance circuit, with a center frequency $f_0 \sim 630$ kHz and bandwidth $\Delta f = 10$ to 30 kHz, then amplified with a cold amplifier (cooled to 4.2 K) and a room-temperature amplifier, to be measured with a spectrum analyzer. Measurements were conducted at three different electron temperatures—10, 21, and 28 mK—and at different propagation lengths (between source and aplifier contacts): 28, 38, 48, and 58 μm.

The general strategy of the measurement is to place the lower-gated half-plane at the tested filling factor ν_{LG}, whereas the upper

half-plane is always tuned to an integer filling, $\nu_{UG} = 0, 1, 2, 3$. The defined chirality of the resultant interface edge modes is always with respect to the chirality of the tested state. For example, for $\nu_{LG} > \nu_{UG}$ the interface charge chirality is DS, whereas for $\nu_{LG} < \nu_{UG}$, the interface charge chirality is US. This is clearly noted in Figs. 1 to 3 (35). A description in terms of one-dimensional (1D) edge channels is a simplification, but one that is useful for illustrations and justified by the topological nature of the relevant phases. To measure the noise excited by the hotspot (at the back of the source) by a single amplifier, the magnetic
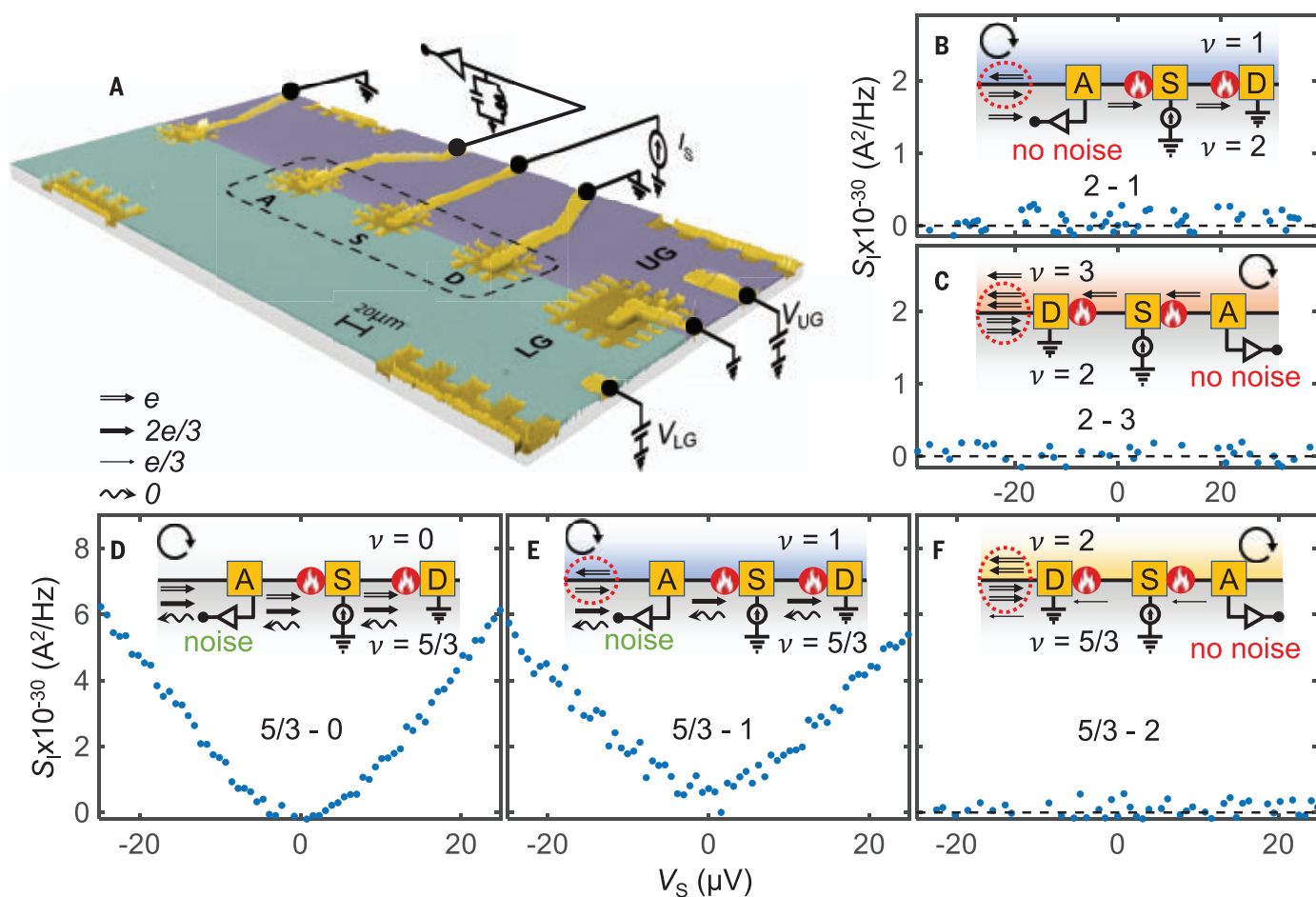
Fig. 1. Experimental setup to create and probe the interface between different states. (**A**) False-colors scanning electron microscopy image of a typical device. Ohmic contacts are in yellow, lower gte (LG) is light blue, and upper gate (UG) is purple. The two-dimensional electron gas (2DEG) is buried 200 nm below the surface [a detailed structure is available in (35)]. Gate voltages, V_{LG} and V_{UG}, control the density. The interface between the two planes hosts interface modes. Ohmic contacts at the edge probe the bulks' filling-factor. Ohmic contacts at the interface probe the interface modes. The heart of the device, highlighted by the dashed box, contains the source contact, S, placed at the same distance from the amplifier contact, A, and the drain contact, D, at different distances S–A: L = 28, 38, 48, and 58 μm. "Blocking contacts" avoid noise arriving at A from secondary hotspots (say, in the drain). (**B** and **C**) Interfaces of integer states ν = 2 with ν = 1 and ν = 3 at 10 mK. Counterpropagating integer modes at the interface are compensated,

leaving a single integer mode at the interface, DS at the 2-1 interface, and US for the 2-3 interface. Hotspots (indicated with "red fire" symbols) are shown on the US side (2-1) and the DS side (2-3) of the source. Noise is not expected. (**D** to **F**) Interfacing the ν = 5/3 state at 10 mK. Arrows indicate four different types of edge modes: Double-line arrow indicates an integer-mode, thick-single arrow indicates a 2/3 charge mode, thin-single arrow indicates a 1/3 charge mode, and waved-arrow indicates a neutral mode. The incomplete circle with an arrowhead indicates the chirality. In (B) to (F), the schematic representation shows nonequilibrated to equilibrated modes from left to right. The 5/3-0 and the 5/3-1 interfaces are presented in the particle-like picture: DS integer 1 and 2/3 modes and US neutral mode. The 5/3-2 is presented in a hole-like picture: two DS integers 1 modes and a US 1/3 mode. The US neutral mode leads to noise in (D) and (E). In (F), a single US 1/3 mode remains, without any observed noise.

Fig. 2. Interfacing abelian fractional states in second Landau level. (**A**) Two-terminal resistance measured at the interface between ν = 2 (or 3) and ν = 7/3, 5/2, and 8/3. (Left) Upper plane is fixed at ν = 2 and (right) upper plane is fixed at ν = 3, whereas the lower plane is swept from ν = 7/3 to ν = 8/3. Clear quantized plateaus corresponding to (left) ν = 1/3, 1/2, and 2/3 and (right) ν = 2/3, 1/2, and 1/3, accurate to ~1%, are observed. Peaks and valleys in between plateaus are caused by reentrant filling factors. (**B**) Interface between ν = 7/3 and ν = 2 at 21 mK. The two integer modes are compensated, leaving the DS 1/3 charge mode. No noise is observed. (**C**) Interface between ν = 7/3 and ν = 3 at 21 mK. Two integer modes are compensated, leaving one US integer and a downstream 1/3. The equilibration of these two counterpropagating charge modes gives rise to an upstream ν = 2/3 charge mode and a downstream neutral mode, accompanied by noise. (**D**) Interface between ν = 8/3 and ν = 2 at 21 mK. Two integer modes are compensated, leaving a DS ν = 2/3 charge mode accompanied by a US neutral mode. US noise is observed. (**E**) Interface between ν = 8/3 and ν = 3 at 21 mK. The equilibration between the integers at the interface leaves a single US charge mode ν = 1/3, with no noise. The arrows in the top left box of (B) to (E) indicate bosonic edge modes with the indicated two-terminal electric conductance.
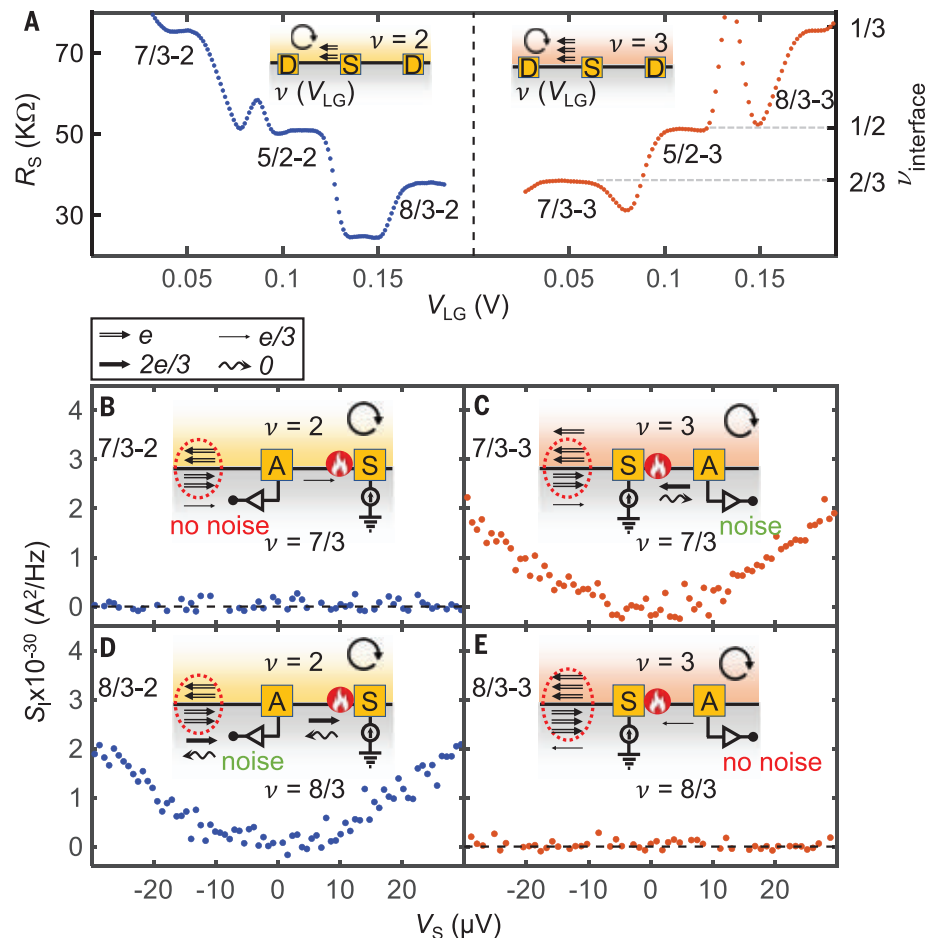
field was reversed (and therefore, so was the chirality) under these two interfacing conditions. In Figs. 1 to 3, for convenience, we flip the amplifier's position instead of the chirality (for example, Fig. 1, B and C).

In Fig. 1, B and C, we show a relatively simple experimental test of interfacing the integer ν = 2 state (tested state) with ν = 1 and ν = 3. The injected DC current, I_S, leads to a hotspot at the US side of the source in Fig. 1B and at the DS side of the source in Fig. 1C, respectively. A perfect charge equilibration took place for all four lengths and three temperatures, with two terminal interface resistance (R_S) = h/e^2; there was no observed US (Fig. 1B) or DS (Fig. 1C) noise. This suggests that the presence of any residual nonequilibrated current (which may persist in spite of charge equilibration) does not lead to observable noise.

Before testing interfacing in the second Landau level, we tested the "interfacing method" with a more complex abelian state in the first Landau level, which involves counterpropagating modes (charge and neutral). We interfaced the ν = 5/3 = 1 + 2/3 filling with the integers ν = 0, 1, 2. We used two similar descriptions of this configuration (fig. S4). The

first assumes that the two half-planes are initially separated, and each of them supports its own edge modes (Fig. 1, E and F). With an intimate proximity at the interface, the integer modes on both sides of the interface compensate each other, leaving only fractional interface modes. In the case of 5/3 interfaces, the integer modes of ν = 1 or ν = 2 are localized, with propagating interface DS ν = 2/3 (with neutral) or US ν = 1/3 modes, respectively (fig. S4, E and G). The second approach is to regard the integer filling $ν_{UG}$ at the upper half-plane as a "vacuum" on which a filling $ν_{LG} − ν_{UG}$ resides in the lower half-plane. Consequently, the common $ν_{UG}$ integer modes circulate around the periphery of the mesa, and the interface carries an edge structure of filling $ν_{LG} − ν_{UG}$ (fig. S4, F and H). We mostly used the first approach.

We returned to the present test of interfacing the 5/3 state (Fig. 1, D to F). Interfacing 5/3-0 or 5/3-1 supports an integer and a fractional ν = 2/3 charge modes or a fractional ν = 2/3 charge mode, respectively, accompanied by an excited US bosonic neutral mode (39–41), leading to the observed US noise (Fig. 1, D and E). Alternatively, interfacing 5/3-2, the com-

pensated integer modes leave behind a single US ν = 1/3 mode at the interface, with no noise observed (Fig. 1F).

We next concentrated on interfacing the dominant fractional states in the second Landau level, ν = 7/3, ν = 5/2, and ν = 8/3 with the integers ν = 2 and ν = 3. Testing first charge equilibration, we fixed the integer filling in the upper half-plane and swept the density of the lower half-plane (Fig. 2A). Clear conductance plateaus, accurate to about 1% (with reentrant peaks and valleys between plateaus), were observed at all propagation lengths and temperatures.

The interfaced 7/3-2 configuration compensates the two integer modes, leaving a DS edge mode of ν = 1/3, with no US noise (Fig. 2B; the chirality is indicated in Fig. 2A). By contrast, the interfaced 7/3-3 configuration leaves the familiar US ν = 2/3 charge mode and a DS bosonic neutral mode (Fig. 2C). The hotspot at the source excites the neutral mode with a DS noise at the amplifier. Interfacing ν = 8/3 with ν = 2 compensates the two integers and leaves a DS charge mode of ν = 2/3 and an US excited bosonic neutral mode (Fig. 2D). By contrast, interfacing the state with ν = 3 leaves
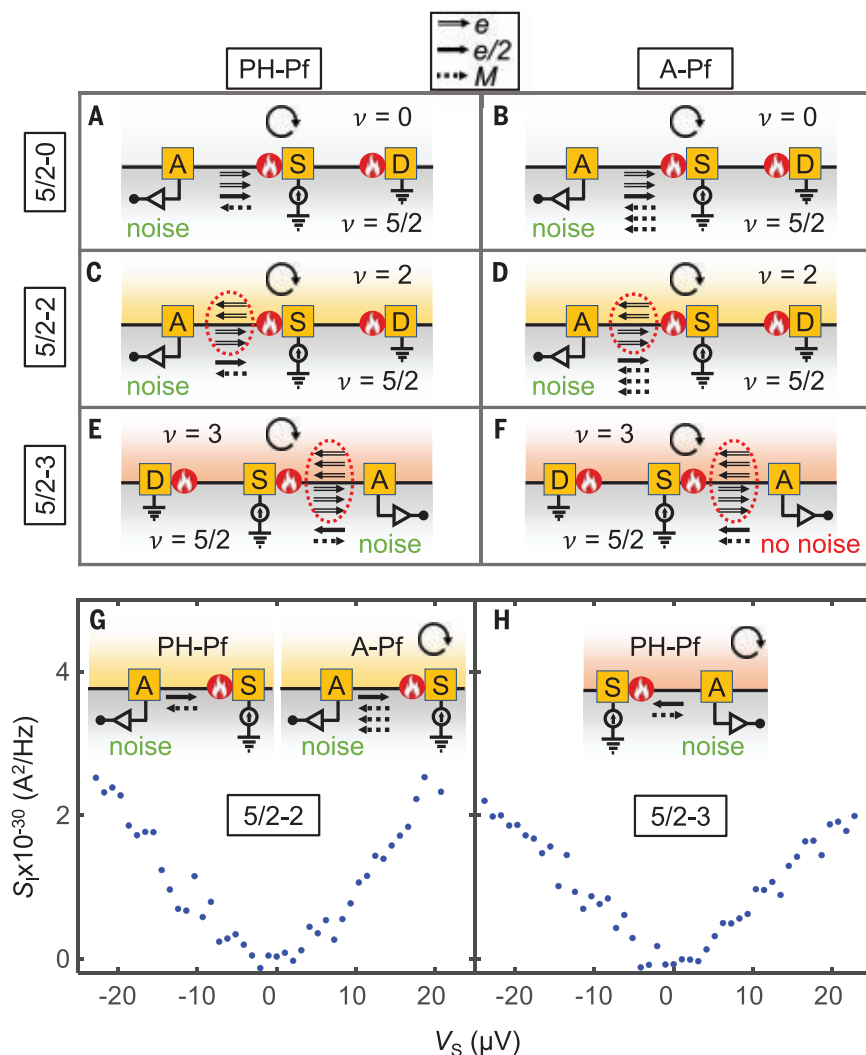
Fig. 3. Interfacing the ν = 5/2 states with ν = 2 and ν = 3. Comparison of edge structures for the PH-Pf and the A-Pf orders with ν = 0, 2, and 3. (**A** and **B**) Edge structure of ν = 5/2 state at the edge of sample presented in particle-like presentation. Because US noise is expected for both orders, this measurement cannot distinguish between the two orders. (**C** and **D**) Interfacing 5/2-2 for both topological orders. US noise is expected for both orders. (**E** and **F**) Interfacing 5/2-3 for both orders, presented for convenience, in the hole-like picture. The three integer modes at the interface are compensated, leaving at the interface PH-Pf, with a US 1/2 charge mode and a DS neutral Majorana mode, and A-Pf, with copropagating US 1/2 charge and neutral Majorana modes. DS noise is expected for the PH-Pf order, whereas no noise is expected for the A-Pf order. (**G**) Measured US noise at the interface of 5/2-2 at 10 mK, which is consistent with both of the competing orders. (**H**) Measured DS noise at the 5/2-3 interface at 10 mK, which is expected only for the PH-Pf order. Double-line arrow indicates an integer-mode, thick-single arrow indicates a 1/2-mode, and the dashed-arrow indicates a Majorana-mode. The incomplete circled arrow indicates the chirality.

a US ν = 1/3 charge mode at the interface, with no resulting noise (Fig. 2E).

Before presenting the main experimental results, it is worth discussing the outcome of interfacing the PH-Pf and A-Pf orders of ν = 5/2 with the integers ν = 0, 2, 3 (Fig. 3, A to F). The consequence of a similar interfacing of other proposed orders of ν = 5/2 are described in figs. S10 and S11 (*35*). The mode structure of the two topological orders of ν = 5/2 with

vacuum—at the 5/2-0 interface—are shown in Fig. 3, A and B. For both orders, the 5/2-2 interface leaves a DS fractional charge mode ν = 1/2 and US Majorana modes, one for the PH-Pf and three for A-Pf orders (Fig. 3, C and D). The 5/2-3 interface is more interesting. An interface of the PH-Pf ν = 5/2 with ν = 3 supports counterpropagating US fractional charge mode ν =1/2 and a DS Majorana mode (Fig. 3E), whereas for the interface of the A-Pf

ν = 5/2 and ν = 3, the latter two modes co-propagate in the US direction (Fig. 3F). Therefore, measuring the chirality of the Majorana mode at the 5/2-3 interface is crucial for identifying the actual topological order of the ν = 5/2 state.

The noise data for 5/2 interfaces are shown in Fig. 3, G and H, measured at 10 mK with a propagation length of 28 μm. Noise was found in the US direction at the interface 5/2-2 (no noise observed in DS) and in the DS direction for 5/2-3 (no US noise was observed), both with similar amplitude at the same source voltage. As discussed above, the measured DS noise at the 5/2-3 interface points at the existence of the PH-Pf order (Fig. 3H, inset). Measurements at all temperatures and lengths (with two different MBE growths and two thermal cycles) led to similar results (figs. S33 and S34). This is the main result of our work.

The PH-Pf is a particle-hole symmetric state. Therefore, the same outcome should occur when it is interfaced with ν = 2, where the system is regarded as a half-filled level of electrons on top of two full Landau levels, and with ν = 3, where the system is regarded as a half-filled Landau level of holes on top of three full Landau levels. Our results manifest this particle-hole symmetry.

The amplitude of the neutral noise as function of the number of compensated integer modes is also important. Given that the temperature of the hotspot (T_{HS}) is proportional to the applied voltage ($T_{HS}^2 \sim KV_S^2$), we plotted the noise data as a function of the source voltage at a fixed propagation length of 28 μm, for a few interfacing conditions (Fig. 4A). The noise (US or DS) was similar for all integers n. The same behavior was also observed for the 8/3-n interfaces (fig. S17B). These results indicate that the integers' hotspots, located at the boundary of the large source ohmic contact, do not take part in the excitation of the neutral (bosonic or Majorana) modes. The latter are excited by the hotspot generated by interedge equilibrations, which are somewhat remote from the source contact (*15*, *29*). The dependence of the noise on the propagation length, for the 5/2-n and 8/3-n interfaces, is shown in Fig. 4B (the solid lines are guides to the eye), indicating a qualitatively similar thermal equilibration process of the "different neutrals" as the propagation length is increased. In fig. S24, we add similar measurements of the interfaces 5/3-n and 2/3-0.

The measured results are naturally explained in terms of interfaces between a PH-Pf topological order of ν = 5/2 and the integers ν = 2 and ν = 3. However, we cannot exclude the possibility of edge reconstruction that might give rise to noise at the interface. In particular, this may include a scenario in which the presence of an interface to an integer filling makes
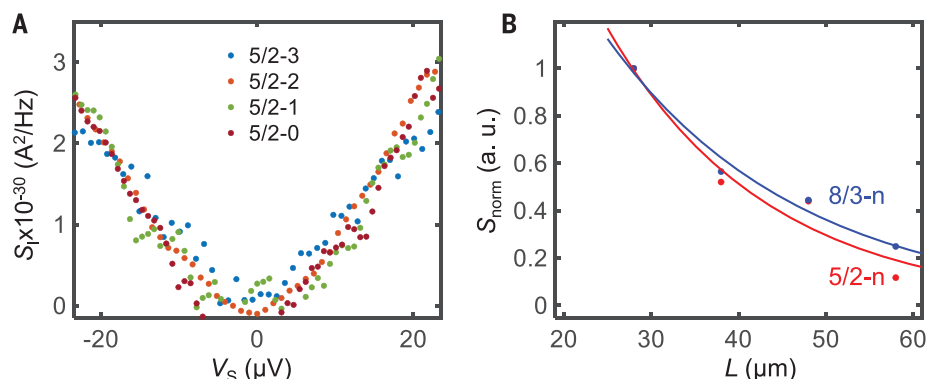
Fig. 4. Dependence of the noise owing to neutral modes on the number of integer modes and the propagation length. (**A**) Noise in a few interfacing conditions, 5/2-n with n = 0, 1, 2, and 3, measured at 10 mK and for 28 μm propagation length. The measured US noise (and DS noise for 5/2-3) is independent of n, indicating that the integer modes do not play a role in the excitation of the neutral modes. A similar observation is found also for 8/3-n, with n = 0, 1, and 2 (fig. S17B). (**B**) The measured noise as a function of distance (normalized to the noise at 28 μm). The solid lines are guides to the eye. Because the noise amplitude does not depend on the number of integer modes [as seen in (A)], the data for ν = 5/2 and ν = 8/3 are annotated as 5/2-n and 8/3-n. The noise decay length is qualitatively similar for both 5/2-n and 8/3-n.

the topological order of the ν = 5/2 close to the interface different from that of the bulk. A specific example of an interface reconstruction in the A-Pf bulk is discussed in (*35*). However, there is no indication for such edge reconstruction in the whole set of the measurements we performed on interfaces of the abelian cases. A thermal metal (*25–28*) at nonzero temperatures may exhibit similar edge physics as that of PH-Pf (*27*) while having small but nonzero bulk longitudinal thermal conductance. This scenario cannot be ruled out by these measurements.

In this work, we have introduced a method that is instrumental in identifying the topological order of the non-abelian ν = 5/2 state. Because the previous experimental determination of the PH-Pf order was based on full thermal equilibration of all modes (*12, 14*), questions were raised as to whether this condition was fulfilled (*15, 30*). Here, by forming chiral 1D modes at the interface between two half-planes, each with a different filling factor, leading to a single ν = 1/2 charge mode and Majorana modes, we considerably strengthen the case for a PH-Pf topological order of the ν = 5/2 state and weaken that of its competitors (*6, 35, 42–44*).

In a broader perspective, a similar interfacing method between quantum states can be used to engineer distinct fractional 1D interface modes, which do not live on the physical edge of the sample. Our method may enable the study of the rich world of non-abelian quasiparticles, including as yet unexplored states such as the theoretically proposed non-abelian ν = 12/5 state.

REFERENCES AND NOTES

1. B. I. Halperin, J. K. Jain, *Fractional Quantum Hall Effects: New Developments*. B. I. Halperin, J. K. Jain, Eds. (World Scientific, 2020).
2. X.-G. Wen, *Quantum Field Theory of Many-Body Systems: From the Origin of Sound to an Origin of Light and Electrons* (Oxford Univ. Press, 2004).
3. B. I. Halperin, *Phys. Rev. Lett.* **52**, 1583–1586 (1984).
4. D. Arovas, J. R. Schrieffer, F. Wilczek, *Phys. Rev. Lett.* **53**, 722–723 (1984).
5. J. M. Leinaas, J. Myrheim, *Nuovo Cimento B* **37**, 1–23 (1977).
6. X. G. Wen, *Phys. Rev. Lett.* **66**, 802–805 (1991).
7. J. B. Pendry, *J. Phys. Math. Gen.* **16**, 2161–2171 (1983).
8. C. L. Kane, M. P. A. Fisher, *Phys. Rev. B Condens. Matter* **55**, 15832–15837 (1997).
9. S. Jezouin *et al.*, *Science* **342**, 601–604 (2013).
10. K. Schwab, E. A. Henriksen, J. M. Worlock, M. L. Roukes, *Nature* **404**, 974–977 (2000).
11. M. Meschke, W. Guichard, J. P. Pekola, *Nature* **444**, 187–190 (2006).
12. M. Banerjee *et al.*, *Nature* **545**, 75–79 (2017).
13. N. Read, D. Green, *Phys. Rev. B Condens. Matter* **61**, 10267–10297 (2000).
14. M. Banerjee *et al.*, *Nature* **559**, 205–210 (2018).
15. K. K. W. Ma, D. E. Feldman, *Phys. Rev. B* **99**, 085309 (2019).
16. K. K. W. Ma, D. E. Feldman, *Phys. Rev. Lett.* **125**, 016801 (2020).
17. A. Aharon-Steinberg, Y. Oreg, A. Stern, *Phys. Rev. B* **99**, 041302 (2019).
18. G. Moore, N. Read, *Nucl. Phys. B* **360**, 362–396 (1991).
19. A. Bid *et al.*, *Nature* **466**, 585–590 (2010).
20. M. Levin, B. I. Halperin, B. Rosenow, *Phys. Rev. Lett.* **99**, 236806 (2007).
21. S. S. Lee, S. Ryu, C. Nayak, M. P. Fisher, *Phys. Rev. Lett.* **99**, 236807 (2007).
22. M. P. Zaletel, R. S. K. Mong, F. Pollmann, E. H. Rezayi, *Phys. Rev. B Condens. Matter Mater. Phys.* **91**, 045115 (2015).
23. E. H. Rezayi, *Phys. Rev. Lett.* **119**, 026801 (2017).
24. P. T. Zucker, D. E. Feldman, *Phys. Rev. Lett.* **117**, 096802 (2016).
25. D. F. Mross, Y. Oreg, A. Stern, G. Margalit, M. Heiblum, *Phys. Rev. Lett.* **121**, 026801 (2018).
26. C. Wang, A. Vishwanath, B. I. Halperin, *Phys. Rev. B* **98**, 045112 (2018).
27. I. C. Fulga, Y. Oreg, A. D. Mirlin, A. Stern, D. F. Mross, *Phys. Rev. Lett.* **125**, 236802 (2020).
28. B. Lian, J. Wang, *Phys. Rev. B* **97**, 165124 (2018).
29. J. Park, C. Spånslätt, Y. Gefen, A. D. Mirlin, *Phys. Rev. Lett.* **125**, 157702 (2020).
30. S. H. Simon, B. Rosenow, *Phys. Rev. Lett.* **124**, 126801 (2020).
31. S. H. Simon, *Phys. Rev. B* **97**, 121406 (2018).
32. D. E. Feldman, *Phys. Rev. B* **98**, 167401 (2018).
33. R. Bhattacharyya, M. Banerjee, M. Heiblum, D. Mahalu, V. Umansky, *Phys. Rev. Lett.* **122**, 246801 (2019).
34. S. K. Srivastav *et al.*, *Phys. Rev. Lett.* **126**, 216803 (2021).
35. Materials and methods are available as supplementary materials.
36. U. Klaß, W. Dietsche, K. Vonklitzing, K. Ploog, *Z. Phys. B Condens. Matter* **82**, 351–354 (1991).
37. J. Park, A. D. Mirlin, B. Rosenow, Y. Gefen, *Phys. Rev. B* **99**, 161302 (2019).
38. C. Spånslätt, J. Park, Y. Gefen, A. D. Mirlin, *Phys. Rev. Lett.* **123**, 137701 (2019).
39. C. L. Kane, M. P. Fisher, J. Polchinski, *Phys. Rev. Lett.* **72**, 4129–4132 (1994).
40. J. Wang, Y. Meir, Y. Gefen, *Phys. Rev. Lett.* **111**, 246803 (2013).
41. Y. Meir, *Phys. Rev. Lett.* **72**, 2624–2627 (1994).
42. G. Yang, D. E. Feldman, *Phys. Rev. B Condens. Matter Mater. Phys.* **90**, 161306 (2014).
43. B. Overbosch, X.-G. Wen, Phase transitions on the edge of the ν = 5/2 Pfaffian and anti-Pfaffian quantum Hall state. arXiv:0804.2087 [cond-mat.mes-hall] (2008).
44. B. I. Halperin, *Helv. Phys. Acta* **56**, 75–102 (1983).
45. B. Dutta *et al.*, Data for: Distinguishing between non-abelian topological orders in a quantum Hall system, Zenodo (2021); doi: 10.5281/zenodo.5638894.

MOLECULAR MAGNETS

Ultrahard magnetism from mixed-valence dilanthanide complexes with metal-metal bonding

Colin A. Gould[1]†, K. Randall McClain[2]†, Daniel Reta[3]‡, Jon G. C. Kragskow[3], David A. Marchiori[4], Ella Lachman[5,6], Eun-Sang Choi[7], James G. Analytis[5,6], R. David Britt[4], Nicholas F. Chilton[3]*, Benjamin G. Harvey[2]*, Jeffrey R. Long[1,6,8]*

Metal-metal bonding interactions can engender outstanding magnetic properties in bulk materials and molecules, and examples abound for the transition metals. Extending this paradigm to the lanthanides, herein we report mixed-valence dilanthanide complexes $(Cp^{iPr5})_2Ln_2I_3$ (Ln is Gd, Tb, or Dy; Cp^{iPr5}, pentaisopropylcyclopentadienyl), which feature a singly occupied lanthanide-lanthanide σ-bonding orbital of $5d_{z^2}$ parentage, as determined by structural, spectroscopic, and computational analyses. Valence delocalization, wherein the d electron is equally shared by the two lanthanide centers, imparts strong parallel alignment of the σ-bonding and f electrons on both lanthanides according to Hund's rules. The combination of a well-isolated high-spin ground state and large magnetic anisotropy in $(Cp^{iPr5})_2Dy_2I_3$ gives rise to an enormous coercive magnetic field with a lower bound of 14 tesla at temperatures as high as 60 kelvin.

Metal-metal bonding underpins a wide range of phenomena in natural and synthetic systems. For example, in the permanent magnets $Nd_2Fe_{14}B$ and $SmCo_5$, strong interactions between the 5d conduction band and localized lanthanide 4f electrons give rise to a large overall moment that couples with the transition metal 3d band to generate large magnetocrystalline anisotropies at high temperature (1). Metal-metal bonding has also been leveraged to make advances in numerous other areas, including multifunctional materials (2), catalysis (3), molecular electronics (4), and molecular magnetism. Of particular relevance here, metal-metal bonding has been used in the design of single-molecule magnets (5, 6), a class of molecules that exhibit intrinsic magnetic bistability and magnetic hysteresis at low temperature (7).

Metal-metal bonding necessarily requires diffuse valence orbitals, and thus nearly all molecular examples involve transition metals with diffuse d orbitals (8). By contrast, the limited radial extension of the lanthanide valence 4f orbitals has thus far precluded isola-

tion of a coordination compound featuring lanthanide-lanthanide bonding. Notably, in endohedral Ln_2 dimers, the transfer of four or five electrons from the lanthanide (Ln) metals to the fullerene cage has been proposed to result in the formation of a doubly or singly occupied σ-bonding orbital of Ln 6s orbital parentage (9). However, Coulombic repulsion between the lanthanide centers is substantially stronger than the bonding interaction, and, as such, these exotic systems are not readily translated to the realm of synthetic molecular chemistry. The isolation of a molecular compound featuring lanthanide-lanthanide bonding would thus represent an important fundamental advance and, furthermore, could enable

access to favorable electronic and magnetic properties, given the large magnetic moments and single-ion magnetic anisotropies of the 4f elements. The recent discovery of divalent lanthanide ions with $4f^n5d^1$ electron configurations (10) suggested to us that it might be possible to achieve lanthanide-lanthanide bonding in a dilanthanide complex with symmetry-compatible 5d orbitals.

Here, we report the mixed-valence dilanthanide complexes $(Cp^{iPr5})_2Ln_2I_3$ (1-Ln; Cp^{iPr5}, pentaisopropylcyclopentadienyl; Ln is Y, Gd, Tb, or Dy), synthesized via reduction of the trivalent precursor complexes $(Cp^{iPr5})_2Ln_2I_4$ with potassium graphite in n-hexane (Fig. 1A). The precursor species were prepared by a salt metathesis reaction between anhydrous LnI_3 and $NaCp^{iPr5}$ [supplementary materials (SM) section 1.2], with the sterically hindered Cp^{iPr5} ligand being selected to favor the formation of a dinuclear complex over higher nuclearity clusters (11). Structural, spectroscopic, and computational analyses of 1-Ln reveal that upon reduction, rather than adopting a configuration with discrete $4f^n5d^1$ Ln^{II} and $4f^n$ Ln^{III} ions, the compounds instead exhibit valence delocalization due to the formation of a singly occupied σ-bonding molecular orbital of d_{z^2} parentage. This bonding interaction corresponds to a Robin-Day Class III formalism (12), wherein the d electron is equally shared by the two lanthanide centers. Valence delocalization imparts strong parallel alignment of the σ-bonding and f electrons on both lanthanides according to Hund's rules, giving rise to high-spin ground states that are thermally isolated even at room temperature (Fig. 1B) (13). In 1-Tb and 1-Dy, the combination
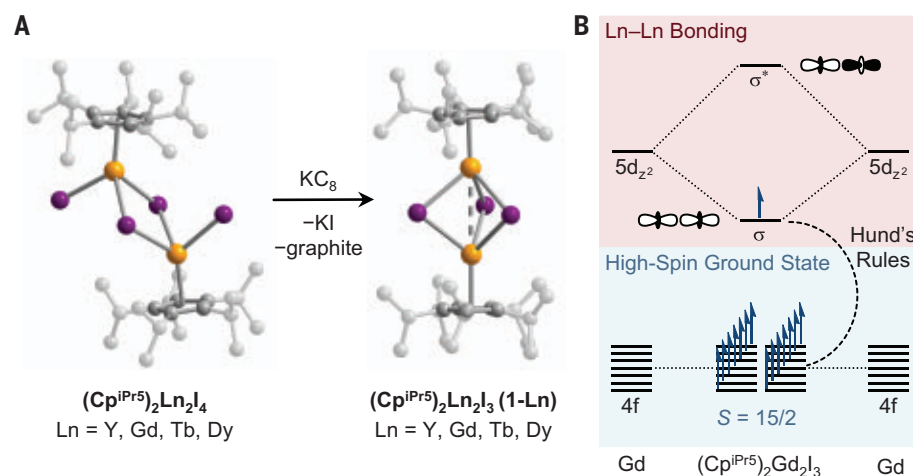
[1]Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA. [2]US Navy, Naval Air Warfare Center, Weapons Division, Research Department, Chemistry Division, China Lake, CA 93555, USA. [3]Department of Chemistry, School of Natural Sciences, The University of Manchester, Manchester M13 9 PL, UK. [4]Department of Chemistry, University of California, Davis, Davis, CA 95616, USA. [5]Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA. [6]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [7]National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32310, USA. [8]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, Berkeley, CA 94720, USA.
*Corresponding author. Email: jrlong@berkeley.edu (J.R.L.); benjamin.g.harvey.civ@us.navy.mil (B.G.H.); nicholas.chilton@manchester.ac.uk (N.F.C.)
†These authors contributed equally to this work.
‡Present address: Kimika Fakultatea, Euskal Herriko Unibertsitatea (UPV/EHU) and Donostia International Physics Center (DIPC), P. K. 1072, 20080 Donostia, Euskadi, Spain.

Fig. 1. Synthesis, x-ray diffraction structures, and molecular orbital diagram of dilanthanide complexes.
(**A**) Single-electron reduction of $(Cp^{iPr5})_2Ln_2I_4$ (left) with potassium graphite in n-hexane affords the compounds $(Cp^{iPr5})_2Ln_2I_3$ (1-Ln) (right). The crystal structures shown here correspond to the case where Ln is Gd. Orange, purple, and gray spheres represent Gd, I, and C atoms, respectively; H atoms are omitted for clarity. (**B**) Molecular orbital diagram for $(Cp^{iPr5})_2Ln_2I_3$ illustrates the formation of a singly occupied σ-bonding orbital of $5d_{z^2}$ parentage and alignment of the σ and f electrons on both lanthanides according to Hund's rules. Here, electrons are filled for the case where Ln is Gd.

of substantial axial magnetic anisotropy and a well-isolated high-spin ground state results in by far the largest coercive magnetic field yet observed for any molecule or molecule-based material.

The solid-state structures of **1-Ln** were determined from single-crystal x-ray diffraction data (Fig. 1A). In all compounds, three iodide anions bridge two metal centers to yield a Ln_2I_3 core with approximate trigonal symmetry, and each metal center is capped with a $(Cp^{iPr5})^-$ ligand. Compounds **1-Y**, **1-Gd**, **1-Tb**, and **1-Dy** display Ln···Ln distances of 3.727(1), 3.769(1), 3.732(1), and 3.713(1) Å (Fig. 2A), respectively, which are within the sum of covalent radii for each metal atom and correspond to formal shortness ratios ranging from 0.961 to 0.981 (8, 14). These ratios are consistent with those determined for homometallic transition metal complexes with half-order metal-metal bonding interactions (15–17). Notably, the lanthanide ions in **1-Ln** are crystallographically equivalent by a C_2 axis, further suggesting the presence of metal-metal bonding. This observation is supported by complete active space self-consistent field (CASSCF) and density functional theory (DFT) calculations performed on **1-Ln**, which predict the presence of a singly occupied molecular orbital (SOMO) that arises from σ-bonding between the d_{z^2} orbitals of the two lanthanide ions (Fig. 2B; figs. S98, S101, and S104; and table S18). The compounds **1-Ln** possess a similar core structure and symmetry to the well-studied confacial bioctahedral clusters of the type $[M_2X_9]^{n-}$, many of which display metal-metal bonding (18, 19). In these trigonally symmetric clusters, the three μ$_2$ bridging halide ligands bring the transition metal centers into close proximity, facilitating a σ-bonding interaction between d_{z^2} orbitals.

To corroborate this hypothesis, we synthesized the complex **1-Tm** (SM section 1.2), which is expected to exhibit a $4f^{12}/4f^{13}$ configuration given that Tm^{III} is known to undergo reduction from a $4f^n$ to a $4f^{n+1}$ electron configuration (10). Single-crystal x-ray diffraction analysis of **1-Tm** (Fig. 2A) revealed a Tm···Tm distance of 3.960 (1) Å, greater than the sum of covalent radii for two Tm atoms [3.80(10) Å] (14) as expected for two noninteracting ions. Additionally, the two Tm ions in **1-Tm** show substantial differences in the average Tm–C [2.583(4) versus 2.658(5) Å] and Tm–I [2.954(3) versus 3.212(4) Å] distances, indicative of valence localization. We also prepared the complex salt $[(Cp^{iPr5})_2Dy_2I_3][B(C_6F_5)_4]$ (**2-Dy**) via iodide abstraction from $(Cp^{iPr5})_2Dy_2I_4$ using $[H(SiEt_3)_2][B(C_6F_5)_4]$ (SM section 1.2), which features a Dy_2I_3 core analogous to that in **1-Dy** but with a $4f^9/4f^9$ configuration. Both the Dy···Dy distance and Dy–I–Dy angle in **1-Dy** are smaller than in **2-Dy** [3.713(1) Å and 75.68(1)° versus 3.902(1) Å and 79.65(1)°, respectively; Fig. 2A], despite the larger ionic

radius expected for Dy upon reduction. Similar trends have been reported for trihalide-bridged transition metal complexes that display metal-metal bonding (18, 19).

Electron paramagnetic resonance (EPR) spectra collected at X-band frequency (9.38 GHz) for **1-Y** (Fig. 3A, fig. S65, and table S8) reveal an axial g tensor, consistent with the presence of a σ-bonding SOMO. Here, **1-Y** is used as an $S = ½$ analog for **1-Ln**, because yttrium is a well-established diamagnetic surrogate for late lanthanide ions. Ultraviolet-visible–near-infrared (UV-Vis-NIR) spectra collected for solutions of **1-Y**, **1-Gd**, **1-Tb**, and **1-Dy** in hexanes (Fig. 3B, left) exhibit intense NIR features [molar extinction coefficient (ε) ≥ 5900 M^{-1} cm^{-1}] at the frequency of maximum absorption ($v_{max} \approx$ 15,000 cm^{-1}, which we assign as intervalence charge transfer (IVCT) bands. Similar features are apparent in diffuse reflectance spectra (Fig. 3B, right). By contrast, strong absorption features are absent in the visible and NIR regions of spectra obtained for **1-Tm**, both as a crystalline solid and in hexanes solution (Fig. 3B). CASSCF multiconfigurational pair-density functional theory (20) and DFT calculations performed on **1-Y** (SM section 8.3) predict that the IVCT band corresponds to excitation from the symmetric d_{z^2}–d_{z^2} σ-bonding SOMO to the antisymmetric d_{z^2}–d_{z^2} σ-antibonding orbital (fig. S98 and tables S18 and S19), and these calculations enabled identification of additional features in the experimental spectrum due to excitations from σ-bonding to δ- and π-type orbitals (inset of fig. S96). The lowest-energy states that approach localized valence are the δ and δ* states, which are substantially higher in energy than the ground state (8600 to 10,300 cm^{-1} higher in energy, respectively). The full-width-at-half-maximum bandwidths of the IVCT features for **1-Y**, **1-Gd**, **1-Tb**, and **1-Dy** are $\Delta v_{1/2}$ = 2370, 2531, 2624, and 2919 cm^{-1}, respectively, considerably narrower than the theoretical bandwidths of $v^o_{1/2}$ = 5753, 5863, 5916, and 5732 cm^{-1}, respectively

(21). The parameter $\Gamma = 1 - \Delta v_{1/2}/v^o_{1/2}$ has been proposed for classifying mixed-valence compounds (22, 23), with $\Gamma > 0.5$ associated with Class III systems. For compounds **1-Y**, **1-Gd**, and **1-Tb**, $\Gamma > 0.5$, whereas for **1-Dy**, $\Gamma \approx 0.5$, consistent with valence delocalization. We note that the well-studied Class III mixed-valence complex $[(tmtacn)_2Fe_2(OH)_3]^{2+}$ (tmtacn, 1,4,7-trimethyl-1,4,7-triazacyclononane), which possesses an M_2X_3 core geometry analogous to that in **1-Ln** and a d_{z^2}–d_{z^2} σ-bonding interaction, exhibits an IVCT feature at 13,500 cm^{-1} with $\Gamma \approx 0.5$ (24).

Variable-temperature, zero-field cooled dc magnetic susceptibility data obtained for **1-Gd**, **1-Tb**, and **1-Dy** under an applied magnetic field of H_{dc} = 1000 Oe reveal a high-spin ground state in each complex that remains thermally well isolated even up to room temperature (Fig. 4A, open symbols). Indeed, at 300 K, the product of the experimental molar magnetic susceptibility and temperature ($\chi_M T$) for **1-Gd**, **1-Tb**, and **1-Dy** is 30.74, 41.91, and 51.77 electromagnetic units (emu) K mol^{-1}, respectively, substantially greater than the values predicted for noninteracting Ln^{II} and Ln^{III} ions (17.88, 26.24, and 31.18 emu K mol^{-1}, respectively) and instead much closer to the theoretical values of 31.88, 48.74, and 58.68 emu K mol^{-1}, respectively, predicted for parallel alignment of the lanthanide $4f^n$ electrons with the unpaired electron in the σ-bonding orbital. The lower magnitude of the experimental $\chi_M T$ values for **1-Tb** and **1-Dy** relative to the theoretical values is associated with the large magnetic anisotropy, as has been observed previously for strongly coupled lanthanide single-molecule magnets (25). A fit to the data for isotropic **1-Gd** (Fig. 4A) using a Heisenberg Hamiltonian for a symmetric exchange-coupled three-spin system (SM section 7.2) gave an exchange constant of J = +387(4) cm^{-1} ($-2J$ formalism, representing spin-spin exchange between the 4f and σ electrons). The exchange constant for **1-Gd** is the largest yet reported
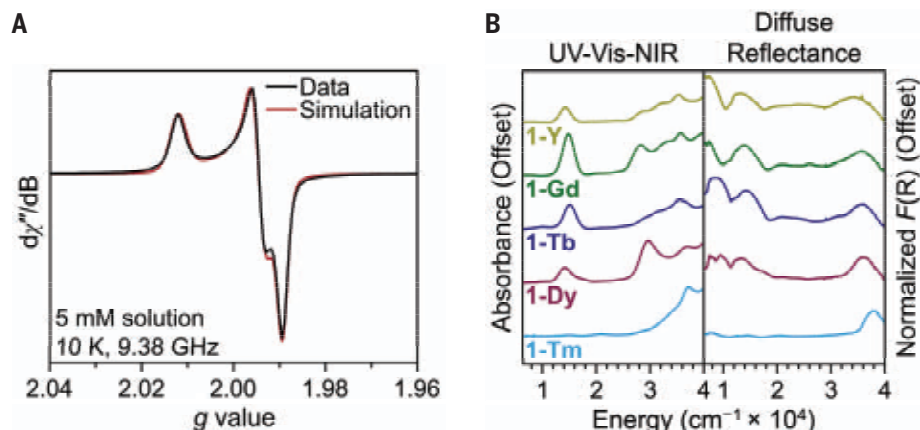

Fig. 2. Structural evidence for lanthanide-lanthanide bonding in $(Cp^{iPr5})_2Ln_2I_3$ (Ln is Y, Gd, Tb, or Dy). (**A**) Comparison of crystal structures of **1-Dy** (left), **2-Dy** (middle), and **1-Tm** (right); green, blue, purple, and gray spheres represent Dy, Tm, I, and C atoms, respectively; isopropyl groups are omitted for clarity. (**B**) SOMO of **1-Gd** as determined by CASSCF calculations; H atoms are omitted for clarity.

Fig. 3. Characterization of valence delocalization. (**A**) Experimental and simulated EPR data for **1-Y**. (**B**) UV-Vis-NIR and diffuse reflectance spectra for **1-Ln**. The bandwidth of the IVCT transition ($\nu_{max} \approx$ 15,000 cm^{-1}) in **1-Y**, **1-Gd**, **1-Tb**, and **1-Dy** supports valence delocalization. $F(R)$ is the Kubelka-Munk transformation of the raw diffuse reflectance spectrum.

for any lanthanide compound, more than twice the value of $J = +170(10)$ cm^{-1} determined for $Gd_2@C_{79}N$ (9). Density matrix renormalization group CASSCF calculations (26) on **1-Gd** (SM section 8.4) revealed a Heisenberg-like spin ladder (table S20) and a value of $J = +389$ cm^{-1}, in excellent agreement with the experimental fit value (fig. S100), along with $J' = -9$ cm^{-1} (4f–4f exchange). Further, broken-symmetry DFT calculations show increasing agreement to experimental values with an increasing fraction of exact exchange (fig. S100 and table S23). Hence, both methods suggest that the exchange coupling is dominated by the direct exchange terms between the σ and 4f electrons (SM section 8.4); thus, the 4f–σ interaction tends to obey Hund's rules.

Field-cooled dc susceptibility data collected for **1-Tb** and **1-Dy** under $H_{dc} = 1000$ Oe diverge from zero-field cooled data at 70 and 75 K, respectively, indicative of magnetic blocking below these temperatures (Fig. 4A), that is, the pinning of the molecular magnetic moments due to the presence of an energy barrier to relaxation. From dc and ac magnetic relaxation data, we extracted 100-s blocking temperatures of $T_b = 65$ and 72 K and large thermal barriers to magnetic relaxation of $U_{eff} = 1383(45)$ and 1631(25) cm^{-1} for **1-Tb** and **1-Dy**, respectively (Fig. 4B). Similar data were obtained for dilute solutions of **1-Tb** and **1-Dy**, confirming the molecular origin of the slow magnetic relaxation (tables S13 and S15). The values for **1-Dy** are the highest reported for any single-molecule magnet, surpassing the previous records of $T_b = 65$ K and $U_{eff} = 1541(11)$ cm^{-1} for $[(Cp^*)Dy(Cp^{iPr5})][B(C_6F_5)_4]$ (27) (Cp^*, pentamethylcyclopentadienyl) (27).

CASSCF calculations were performed on model systems $[(Cp^{iPr5})_2LnLuI_3]^+$ ($4f^n/4f^{14}$) and $(Cp^{iPr5})_2LnLuI_3$ ($4f^n$–σ–$4f^{14}$) for the cases where Ln is Dy or Tb to elucidate changes to the local magnetic anisotropy due to the

presence of the σ-bonding electron and quantify the $4f^n$–σ coupling (SM sections 8.5 and 8.6). The calculated crystal field splitting in $[(Cp^{iPr5})_2DyLuI_3]^+$ is similar to that of the isolated Dy^{III} ions in **2-Dy** (table S33, compare with table S40), whereas the crystal field splitting in $(Cp^{iPr5})_2DyLuI_3$ is substantially enhanced by the σ-bonding electron (table S36); the same is true for $(Cp^{iPr5})_2TbLuI_3$ (table S29, compare with table S26). By generalizing an exchange coupling model proposed for orbitally degenerate ions (SM section 8.5) (28), we calculated that the magnetic coupling in $(Cp^{iPr5})_2DyLuI_3$ and $(Cp^{iPr5})_2TbLuI_3$ is dominated by isotropic spin-spin Heisenberg exchange, with $J = +524$ and +519 cm^{-1}, respectively ($-2J$ formalism). This exchange is stronger than the 4f spin-orbit coupling, and hence the σ electron spin first couples to the total spin of the 4f shell, followed by coupling to the orbital angular momentum. These results are consistent with those for **1-Gd**, revealing that the exchange interactions are dominated by direct exchange, although substantial anisotropy arises in orbitally dependent exchange terms for $(Cp^{iPr5})_2DyLuI_3$ and $(Cp^{iPr5})_2TbLuI_3$ that are absent in **1-Gd**. We next built a model Hamiltonian to describe the full exchange spectrum for **1-Tb** (Fig. 4C); this analysis is more complicated for **1-Dy** owing to the proximity of the first-excited ^6F term to the ground ^6H term of Dy^{III} (SM section 8.6). The ground Kramers doublet for **1-Tb** is $M_J = \pm25/2$, arising from the maximal spin $^{14}I_{25/2}$ term (the ground doublet for **1-Dy** is $M_J = \pm31/2$ from the $^{12}N_{31/2}$ term), although different spin multiplicities are close in energy. Thus, the barrier to magnetic relaxation in **1-Tb** (and, by extension, **1-Dy**) originates from very strong Heisenberg spin-spin coupling between the 4f and σ electrons and large magnetic anisotropy at the lanthanide ion (tables S31 and S39). The larger U_{eff} value for **1-Dy** versus **1-Tb** arises from a combination of a

slightly stronger crystal field as well as considerably stronger orbitally dependent exchange interactions (table S35, compare with table S28).

Variable-field magnetization data obtained between ±14 T for **1-Dy** and **1-Tb** revealed wide magnetic hysteresis loops that remain open to 80 and 64 K, respectively, among the highest temperatures reported for any single-molecule magnet (figs. S92 and S94). For both compounds, the coercive field (H_c) increases dramatically as the temperature is lowered, and it was not possible to saturate the magnetization below 68 K for **1-Dy** or 62 K for **1-Tb** using a conventional magnetometer ($H_{dc,max} = 14$ T). As such, we can only determine a lower bound of $H_c \geq 14$ T below 60 K for **1-Dy** (Fig. 4D) and below 50 K for **1-Tb**. This value far surpasses the previous record of 7.9 T at 10 K for the single-molecule magnet $[(Cp^{Me4H}_2Tb)_2(\mu–N_2^\bullet)]^-$ (25) and coercivities of commercial magnets $SmCo_5$ (4.3 T at 4.2 K) (29) and $Nd_2Fe_{14}B$ (5.0 T at 80 K) (30). Indeed, this lower bound H_c for **1-Dy** and **1-Tb** is the largest yet reported for any molecule or molecule-based material. We can also characterize the magnetic energy density, $(BH)_{max}$, of **1-Dy** and **1-Tb** (SM section 7.7), and these values are more than an order of magnitude greater than those determined for other single-molecule magnets at comparable temperatures and even begin to approach those reported for $SmCo_5$ and $Nd_2Fe_{14}B$ (table S17). Further investigation of the magnetic properties of **1-Tb** and **1-Dy** at higher magnetic fields ($H_{dc} > 14$ T) is ongoing, and preliminary data indicate that the values of H_c and $(BH)_{max}$ for these compounds are substantially underestimated by measurements between ±14 T. Indeed, field-cooled demagnetization data collected for **1-Tb** between ±35 T reveal a lower-bound estimate for the coercive magnetic field of 25 T at 50 K (fig. S93), and even a field of +35 T is unable to fully saturate the magnetization of **1-Tb** at this temperature.

Notably, **1-Dy** and **1-Tb** display $H_c \geq 14$ T at relatively high temperatures, conditions under which the vast majority of single-molecule magnets display closed hysteresis loops. For comparison, only five other molecules—all of the form $[Cp^R_2Dy]^+$—show open hysteresis loops at 60 K, and the next highest H_c value at this temperature is 2.1 T for $[(Cp^*)Dy(Cp^{iPr5})]^+$ (27). The larger H_c values for **1-Tb** and **1-Dy** relative to $[Cp^R_2Dy]^+$ complexes can be attributed to the mitigation of Raman and quantum tunneling processes, which are the predominant pathways for magnetic relaxation in mononuclear single-molecule magnets below the blocking temperature. Indeed, **1-Dy** exhibits a magnetic relaxation time (τ) ~ 1000 s at 67 K, whereas $[(Cp^*)Dy(Cp^{iPr5})]^+$ does not display τ ~ 1000 s until $T \leq 31$ K. Previous studies of dinuclear lanthanide compounds have demonstrated that strong magnetic coupling can
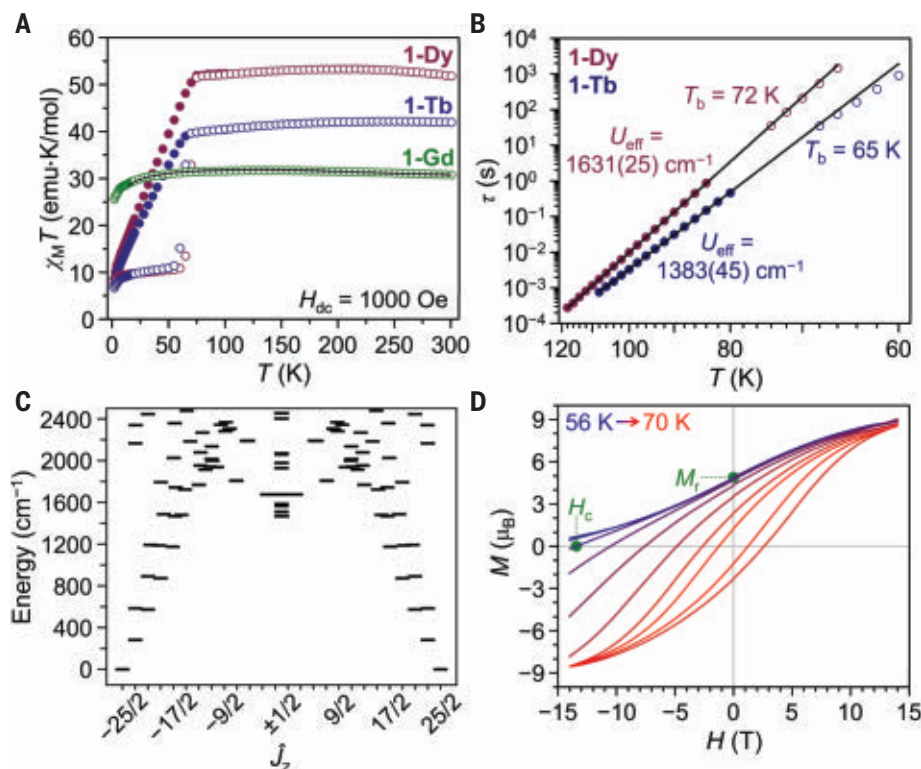
Fig. 4. Magnetic characterization data. (**A**) Field-cooled (filled circles) and zero-field cooled (open circles) dc magnetic susceptibility data for **1-Gd**, **1-Dy**, and **1-Tb**. A fit to the data for **1-Gd** (black line) gives an exchange constant of $J = +387(4)$ cm^{-1}. (**B**) Plots of magnetic relaxation time (τ, log scale) versus T (inverse scale) for **1-Tb** and **1-Dy** derived from ac magnetic susceptibility (filled circles) and dc relaxation (open circles) data; black lines represent fits to the expression $\tau^{-1} = \tau_0^{-1}\exp(-U_{eff}/k_B T)$, yielding U_{eff} and τ_0 values of 1383(45) cm^{-1} and $10^{-11.1(1)}$ s and 1631(25) cm^{-1} and $10^{-12.2(3)}$ s, respectively (k_B, Boltzman constant). (**C**) Ground state splitting in **1-Tb** derived from CASSCF calculations performed on a model (Cp^{iPr5})$_2$TbLuI$_3$ complex, after downscaling the calculated exchange coupling by 20% (SM section 8.5 and fig. S103). (**D**) Field-cooled demagnetization data from 56 to 66 K (open curves; 2K steps) and magnetic hysteresis data at 68 and 70 K for **1-Dy** (closed loops; sweep rate of 100 Oe s^{-1}).

suppress Raman and quantum tunneling processes, resulting in large values of H_c (*25*). However, the exchange coupling for **1-Dy** and **1-Tb** is an order of magnitude greater than that for other dinuclear lanthanide complexes, and the present compounds are the first examples where the exchange interaction is collinear with the large magnetic anisotropy induced by the crystal field. These two features combine to give the very large values of H_c observed for **1-Dy** and **1-Tb**. Further investigation into the magnetic relaxation dynamics of these compounds is ongoing, and we hope that recent perspectives on the theory of magnetic relaxation (*31–33*), new methods for ab initio spin dynamics (*34, 35*), and more detailed experimental studies (*36*) can be brought to bear to elucidate their behavior.

By exploiting the distinctive electronic structure of divalent lanthanides with $4f^n5d^1$ configurations, we have isolated the complexes (Cp^{iPr5})$_2$Ln$_2$I$_3$ (**1-Ln**; Ln is Gd, Tb, or Dy) featuring formal lanthanide-lanthanide bonds. Strong parallel alignment of the $4f^n$ electrons

on each lanthanide with a single electron in the σ-bonding orbital of $5d_{z^2}$ parentage gives rise to thermally well-isolated, high-spin ground states according to Hund's rules, and, in the case of **1-Tb** and **1-Dy**, record coercivities at liquid nitrogen temperatures that surpass even commercial magnets. It is exciting to consider the prospect of designing extended solids in which such lanthanide-lanthanide bonded units are coupled through exchange interactions as a means of creating powerful next-generation permanent magnets.

REFERENCES AND NOTES

1. L. Lanotte, F. Lucari, L. Pareti, *Magnetic Properties of Matter* (World Scientific, 1996).
2. C. Pfleiderer *et al.*, *Nature* **412**, 58–61 (2001).
3. I. G. Powers, C. Uyeda, *ACS Catal.* **7**, 936–958 (2017).
4. J. F. Berry, in *Metal–Metal Bonding*, G. Parkin, Ed., vol. 136 of *Structure and Bonding*, D. M. P. Mingos, Ed. (Springer, 2010).
5. R. Hernández Sánchez, T. A. Betley, *J. Am. Chem. Soc.* **137**, 13949–13956 (2015).
6. K. Chakarawet, P. C. Bunting, J. R. Long, *J. Am. Chem. Soc.* **140**, 2058–2061 (2018).
7. D. Gatteschi, R. Sessoli, J. Villain, *Molecular Nanomagnets* (Oxford Univ. Press, 2006).
8. F. A. Cotton, C. A. Murillo, R. A. Walton, *Multiple Bonds Between Metal Atoms* (Springer, 2005).
9. F. Liu *et al.*, *Acc. Chem. Res.* **52**, 2981–2993 (2019).
10. M. E. Fieser *et al.*, *J. Am. Chem. Soc.* **137**, 369–382 (2015).
11. W. J. Evans *et al.*, *J. Coord. Chem.* **59**, 1069–1087 (2006).
12. M. B. Robin, P. Day, in vol. 10 of *Advances in Inorganic Chemistry and Radiochemistry*, H. J. Emeléus, A. G. Sharpe, Eds. (Academic Press, 1968), pp. 247–422.
13. Here, we use the term "high-spin ground state" to indicate that the ground state for **1-Ln** (Ln is Gd, Tb, or Dy) possesses the highest possible value of spin (*S*) owing to strong parallel alignment of both the 4f and σ-bonding electrons. The compounds **1-Tb** and **1-Dy** also possess unquenched orbital angular momentum (*L*), and thus the full electronic states are best described by the total angular momentum (*J*).
14. B. Cordero *et al.*, *Dalton Trans.* **2008**, 2832–2838 (2008).
15. R. J. Eisenhart, L. J. Clouston, C. C. Lu, *Acc. Chem. Res.* **48**, 2885–2894 (2015).
16. C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward, *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.* **72**, 171–179 (2016).
17. Reference codes YOLYIH, WECDAJ, EHIMAJ, and VADDEJ01 in (*16*).
18. H.-B. Kraatz, P. M. Boorman, *Coord. Chem. Rev.* **143**, 35–69 (1995).
19. F. A. Cotton, D. A. Ucko, *Inorg. Chim. Acta* **6**, 161–172 (1972).
20. G. Li Manni *et al.*, *J. Chem. Theory Comput.* **10**, 3669–3680 (2014).
21. N. S. Hush, *Prog. Inorg. Chem.* **8**, 391–444 (1967).
22. B. S. Brunschwig, C. Creutz, N. Sutin, *Chem. Soc. Rev.* **31**, 168–184 (2002).
23. D. M. D'Alessandro, F. R. Keene, *Chem. Soc. Rev.* **35**, 424–440 (2006).
24. D. R. Gamelin, E. L. Bominaar, M. L. Kirk, K. Wieghardt, E. I. Solomon, *J. Am. Chem. Soc.* **118**, 8085–8097 (1996).
25. S. Demir, M. I. Gonzalez, L. E. Darago, W. J. Evans, J. R. Long, *Nat. Commun.* **8**, 2144 (2017).
26. S. Keller, M. Dolfi, M. Troyer, M. Reiher, *J. Chem. Phys.* **143**, 244118 (2015).
27. F. S. Guo *et al.*, *Science* **362**, 1400–1403 (2018).
28. N. Iwahara, L. F. Chibotaru, *Phys. Rev. B* **91**, 174438 (2015).
29. R. Kütterer, H.-R. Hilzinger, H. Kronmüller, *J. Magn. Magn. Mater.* **4**, 1–7 (1977).
30. C. Benabderrahmane *et al.*, in *Proceedings of the 11th European Particle Accelerator Conference*, I. Andrian, C. Petit-Jean-Genaz, Eds. (EPAC, 2008), pp.WEPC098.
31. L. T. A. Ho, L. F. Chibotaru, *Phys. Rev. B* **97**, 024427 (2018).
32. L. Gu, R. Wu, *Phys. Rev. Lett.* **125**, 117203 (2020).
33. L. Gu, R. Wu, *Phys. Rev. B* **103**, 014401 (2021).
34. D. Reta, J. G. C. Kragskow, N. F. Chilton, *J. Am. Chem. Soc.* **143**, 5943–5950 (2021).
35. M. Briganti *et al.*, *J. Am. Chem. Soc.* **143**, 13633–13645 (2021).
36. A. Chiesa *et al.*, *Phys. Rev. B* **101**, 174402 (2020).
37. C. A. Gould, J. R. Long, Powder x-ray diffraction data, Dryad (2022); https://doi.org/10.6078/D1RT4W.

PLANETARY SCIENCE

Measuring the melting curve of iron at super-Earth core conditions

Richard G. Kraus[1]*, Russell J. Hemley[2], Suzanne J. Ali[1], Jonathan L. Belof[1], Lorin X. Benedict[1], Joel Bernier[1], Dave Braun[1], R. E. Cohen[3], Gilbert W. Collins[4], Federica Coppari[1], Michael P. Desjarlais[5], Dayne Fratanduono[1], Sebastien Hamel[1], Andy Krygier[1], Amy Lazicki[1], James Mcnaney[1], Marius Millot[1], Philip C. Myint[1], Matthew G. Newman[6], James R. Rygg[4], Dane M. Sterbentz[1], Sarah T. Stewart[7], Lars Stixrude[8], Damian C. Swift[1], Chris Wehrenberg[1], Jon H. Eggert[1]

The discovery of more than 4500 extrasolar planets has created a need for modeling their interior structure and dynamics. Given the prominence of iron in planetary interiors, we require accurate and precise physical properties at extreme pressure and temperature. A first-order property of iron is its melting point, which is still debated for the conditions of Earth's interior. We used high-energy lasers at the National Ignition Facility and in situ x-ray diffraction to determine the melting point of iron up to 1000 gigapascals, three times the pressure of Earth's inner core. We used this melting curve to determine the length of dynamo action during core solidification to the hexagonal close-packed (hcp) structure. We find that terrestrial exoplanets with four to six times Earth's mass have the longest dynamos, which provide important shielding against cosmic radiation.

With the discovery of planets outside our Solar System, the search for life on other planetary bodies is one of the grand challenges of our time. This search among the diverse landscape of planets has driven the need for a deeper understanding of the formation and evolution of these extrasolar bodies. The sheer abundance of iron within rocky planet interiors motivates our need for a better understanding of the properties and response of iron at the extreme conditions deep in the cores of more massive Earth-like planets. Specifically, the iron melting curve is critical to understanding internal structure, thermal evolution, and the potential for dynamo-generated magnetospheres—a magnetosphere is believed to be an important component of

habitable terrestrial planets, as it is on Earth (1, 2).

Earth's magnetodynamo is generated in the convecting liquid outer core surrounding the solid inner core. Within the convecting liquid, as a parcel of iron descends toward the center of a planet, it is compressed at constant entropy, implying that it is compressed faster than heat can flow, but not so fast that viscous stresses dissipate substantial energy. The isentropic temperature profile in the liquid iron alloy in the outer core of Earth intersects the iron melting curve at ~330 GPa, forming the outer boundary of the solid inner core (3). The current highest-pressure melting curve data on pure iron reach only 290 GPa (4). The high pressure-temperature (P-T) phase relations also continue to be debated at these pressures despite the numerous studies that have been performed over the years (5–9). As a result of the sensitive interplay between the temperature profile and the melting transition defining the fate of iron-rich cores of rocky planets, a range of disparate predictions exist for magnetospheres of super-Earths. Some predict that only Earth-sized or smaller planets can have a liquid outer core, while anything larger will have a completely solid core (10). The opposite scenario has also been proposed, that is, that planets larger than Earth will have completely liquid cores (11). Another model suggests that planets larger than five Earth masses

($M_\oplus$) can have partially molten liquid outer cores (12). A core evolution model has also been proposed in which the iron isentrope is steeper, in the P-T plane, than the melting curve for bodies larger than a few Earth masses, causing iron crystallites to form at the outer radius of a liquid core and "snow" into the center of the planet, which the authors suggest will inhibit convection and a dynamo (13). This breadth of predicted phenomena is not unexpected given that the melting curve of iron is extrapolated by more than an order of magnitude in pressure for each of these models (11).

We experimentally determined the high-pressure melting curve and structural properties of pure iron up to 1000 GPa, or nearly four times greater pressure than any previous experiments. At the National Ignition Facility (NIF) of Lawrence Livermore National Laboratory (LLNL), we performed a series of seven experiments that emulate the conditions observed by a parcel of iron descending toward the center of a super-Earth core. As illustrated in Fig. 1, we tailored the laser power as a function of time for 16 NIF laser beams incident upon our sample package to first create a single steady shock wave in an iron sample, taking the iron to a state on the Hugoniot between 220 and 300 GPa and setting the entropy of the system. As the shock transits from the iron into a lithium fluoride (LiF) window, the iron decompresses isentropically to pressures between 120 and 160 GPa, ensuring a completely liquid iron sample for all but the lowest shock pressure experiment. Then we precisely increase the laser power to isentropically compress the sample to the desired peak pressure, up to 1000 GPa in ~10 ns, where the initial shock pressure and peak pressure in the sample are accurately determined from the measured interface velocity between the iron sample and LiF window (14). To document the atomic structure of the iron while it is at peak pressure, another 24 beams of the NIF laser illuminate a germanium or zirconium foil, producing a hot plasma. The plasma emits a ~1-ns burst of He-α radiation at 10.25 or 16.25 keV with a bandwidth of ~1%, allowing us to record an x-ray diffraction snapshot of the desired pressure state using image plates within the TARDIS diagnostic (15, 16). A 400-μm-diameter pinhole of palladium or platinum attached to the sample package collimates the x-rays and casts a known diffraction

[1]Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. [2]Departments of Physics, Chemistry, and Earth and Environmental Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA. [3]Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC 20015, USA. [4]Department of Mechanical Engineering, Department of Physics and Astronomy, and Laboratory for Laser Energetics, University of Rochester, Rochester, NY 14627, USA. [5]Sandia National Laboratories, Albuquerque, NM 87123, USA. [6]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA. [7]Department of Earth and Planetary Sciences, University of California Davis, Davis, CA 95616, USA. [8]Department of Earth, Planetary, and Space Sciences, University of California Los Angeles, Los Angeles, CA 90095, USA.
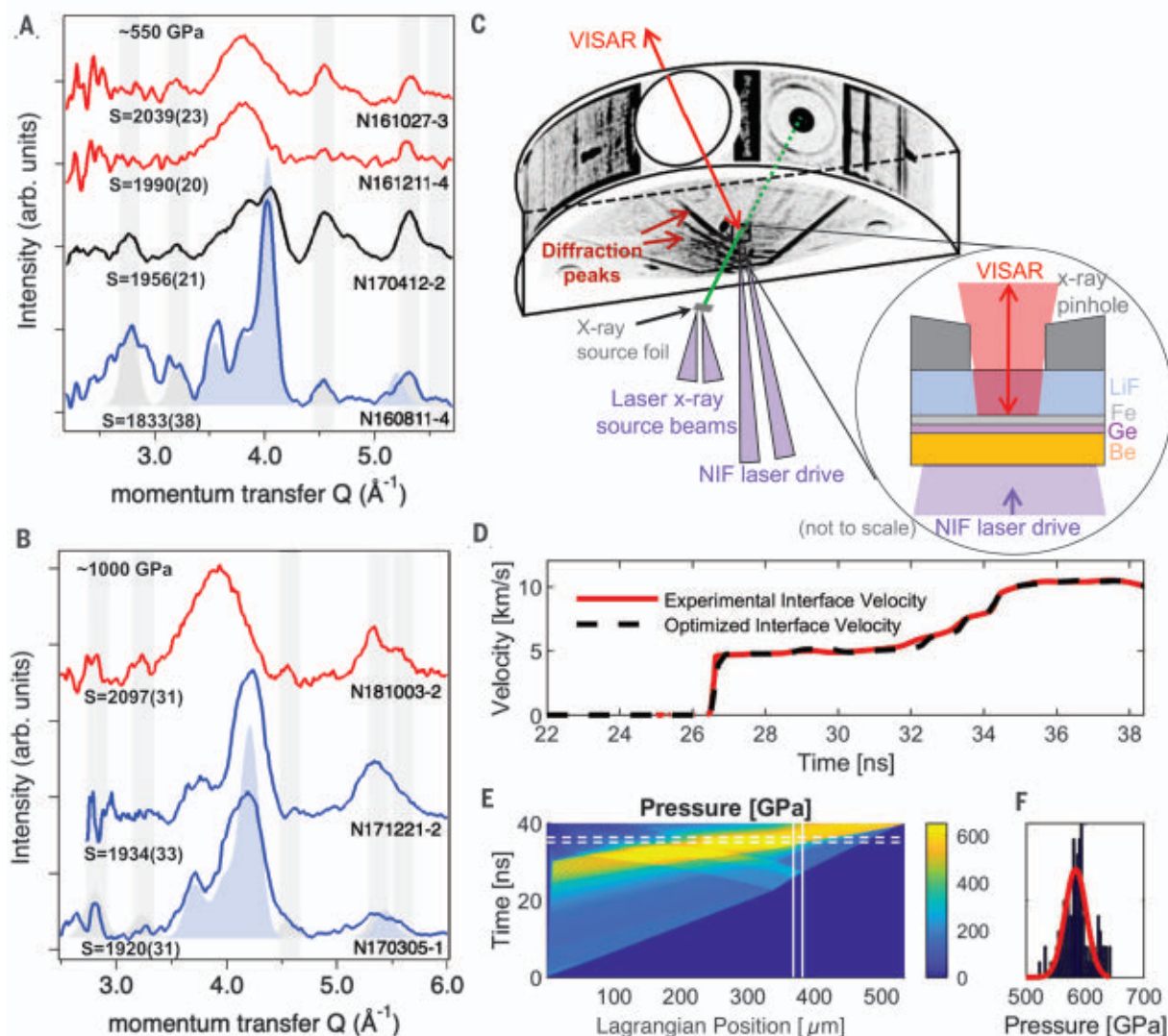*Corresponding author. Email: kraus4@llnl.gov

Fig. 1. Determining the phase assemblage of iron. (**A**) Diffraction lineouts for iron at ~550 GPa, with gray vertical bars marking the positions of diffraction peaks from the pinhole that are used for calibration. Red curves represent liquid iron; black, mixed-phase iron; and blue, hcp iron. The shaded blue area represents the ideal hcp pattern for the diagnostic resolution with a Ge backlighter. Also noted are the entropy states, S, in joules per kilogram per kelvin, and the shot numbers, N. (**B**) As in (A), but at a peak pressure of ~1000 GPa. (**C**) Schematic experimental configuration of the TARDIS diagnostic, with a zoomed-in view of the sample package that is attached to a collimating pinhole at the front of the TARDIS. (**D**) VISAR (velocity interferometer system for any reflector) data from N170412-2, which is used for direct impedance matching to determine the initial shock pressure and as input for a forward optimization (black dashed line). The forward optimization provides a pressure history throughout the sample package (**E**), from which one can infer the pressure history in the sample, defined by the vertical white bars, and (**F**), a pressure histogram of the iron during the x-ray exposure, denoted by the horizontal white dashed bars.

pattern that we use to calibrate the location of the image plates in the TARDIS (*15, 16*).

To constrain the melting curve, we performed sets of experiments at about the same peak pressure with different initial shock strengths, scanning entropy states that bound the melting curve. We present a summary of the in situ x-ray diffraction data (Fig. 1), where we observed pressure-driven solidification of iron from an initially liquid state into the hcp structure on the nanosecond time scale. With decreasing entropy at a given peak pressure, as within a cooling planet, the material state

changes from liquid iron, as we evidenced by purely diffuse x-ray scattering, to a mixed state of hcp and liquid, and finally to solid hcp iron. At ~1000 GPa, corresponding to greater depths in a planet, we again observed the transition from liquid iron at high entropy states to solid hcp iron as the entropy is lowered, albeit at a higher entropy than at the 550 GPa peak pressure experiments.

The observation of pressure-driven solidification indicates that the melt curve is steeper than the isentrope. Our measured experimental bounds on the solidus and liquidus at two

discrete peak pressures further constrain the melt curve to be steeper than the isentrope. These observations reaffirm the expected phenomena of bottom-up core solidification, where dynamo simulations show that the presence of bottom-up solidification will produce stronger magnetic fields than in the alternative case of top-down solidification (*17*). Furthermore, our observation of hcp iron along the melt curve combined with that of Turneaure *et al.* (*8*) refutes predictions of body-centered cubic stability in pure iron at core conditions (*18, 19*), where it is noted

Fig. 2. The melting curve of iron. (**A**) Phase diagram of iron in the P-S plane. The phases we observed (hcp, hcp+liquid, and liquid) and the pressures for incipient and complete melting (21) were used to constrain power-law fits to the high-pressure melt curve (15). (**B**) Comparison of previous melting curve measurements on iron, all below 290 GPa, theoretical estimates, and our melting curve up to 1000 GPa. Transformation of measured melt curve from the P-S plane to P-T using experimentally constrained C_V and γ, with shaded 1σ uncertainties based on the uncertainty in the melt temperature at 260 GPa, Cv, γ, and our measured melt curve in the P-S plane. Also included are the phase measurements as a function of pressure and temperature, where individual 1σ temperature uncertainties are ~800 K.

Fig. 3. Relative time scale for core solidification as a function of planetary mass. The time scale for solidification, τ, with dashed lines representing 1σ uncertainties, is based on the balance of heat flux out of the core, Q_{CMB} (29), with the reduction in entropy that is required to solidify the entire core (15). The time scale for solidification increases with planet mass, with a maximum in the 4 to 6 Earth mass ($M_\oplus$) regime. Our predicted time scale for solidification of Earth's core is 6.2(3.4) Gyr, where the lifetime of our star is ~9 to 10 Gyr and increases for lower-mass stars (15).

that substantial alloying can affect phase stability and the mode of core solidification (20). Finally, the solidification into a mixed phase suggests that the nanosecond time scale of the experiments is not causing the transition to be substantially overdriven, where we might only expect to see a liquid or completely solidified system (14, 15). This observation provides confidence in our measurement of the equilibrium melt curve.

Using available experimental data for the shock temperature and sound speed of iron in the liquid phase (9, 21), we constrained the isochoric heat capacity, C_V = 4.2(1.0) kb per atom, and the Grüneisen parameter, γ = 1.51(5), along the Hugoniot from 6000 to 8000 K and 280 to 320 GPa. (21). The entropy on the Hugoniot is referenced to the high-temperature entropy of iron at 1 bar (22) via an isentrope intersecting the Hugoniot at 260 GPa and 5530 K (15). From analytic expressions for the entropy change along the Hugoniot (23), we obtained the entropy as a function of the shock pressure in each experiment (table S2). The solidus and liquidus are referenced to the pressures for incipient and complete melting

reported in (21) and then fit to the phase measurements obtained in this study, assuming a constant entropy of melting (Fig. 2).

Whereas the pressure-entropy (P-S) plane is the natural thermodynamic space for evaluating changes in phase along an isentrope, such as in volcanism or planetary impacts (24–26), the P-T plane is more common for comparing phase diagrams and assessing the accuracy of theoretical models. To convert our data, melt curve, and uncertainties into the P-T plane, we reevaluate the liquid phase–only shock temperature measurements of Yoo et al. (9). This eliminates complications associated with thermal conductivity and release paths in mixed phases. We then extrapolate the shock temperatures down to 260(3) GPa, the pressure of complete melting along the principal Hugoniot (21), and find the temperature on the melting curve of iron at 260(3) GPa to be 5530(530) K, which is in excellent agreement with the extrapolation of the melting curve of Anzellini et al. (27). From this reference point on the melting curve, our experimentally derived C_V and γ, and our measured curve in the P-S plane, we obtain the temperature along the melting curve

from 260 to 1000 GPa through a two-step thermodynamic integration, with a Simon fit of the melting transition temperature T_m = 5530 $[(P - 260)/293 + 1]^{0.552}$ (Fig. 2) (15).

Our direct measurements provide an experimentally constrained melting curve of iron to nearly four times greater pressure than any previous measurement. Although our data are focused on iron melting in super-Earth cores, our results also provide accurate determination of iron melting at the P-T conditions from the bottom of Earth's core through the inner core boundary (ICB), which has remained controversial because of the lack of direct measurements of melting through this regime. The differences in reported melting behavior of iron at core conditions have narrowed over the years, but the estimated T_m at the ICB at 330 GPa has still been based on extrapolation of data. Because our data provide information on the solidification of the cores of rocky planets well beyond the conditions of Earth's interior, we can interpolate to find the melting temperature of iron at the ICB, where we find T_m = 6230(540) K. This value is similar, within the uncertainties, to that extrapolated from lower-pressure melting data using thermodynamic constraints (5, 6) and recent estimates (4, 7, 8). Comparing first-principles calculations, our results are consistent with, but somewhat shallower than, the density functional theory (DFT) fit (28) that Stixrude reported (29). Our melting curve is substantially steeper than that predicted by other DFT-based calculations (30) and the

Lindemann-based melting curve used in magnetodynamo lifetime calculations (*31*).

Assuming Earth's core mass fraction and compressibility (*12*, *32*), we calculated pressure profiles within the mantles and cores of super-Earths. From this structure calculation and our measurement of the liquidus in the *P-S* plane, we directly determined how much the average entropy of the liquid iron core needs to drop to solidify from the center to the core-mantle boundary (CMB) as a function of planet size. Given that iron cores will likely start as completely liquid owing to the considerable entropy generation by giant impacts during the late stages of accretion (*15*, *33*, *34*) and assuming a CMB heat flux, Q_{CMB}, of $80(M/M_\oplus)$ mW m^{-2} (*29*), we estimate the time scale for solidification of super-Earth cores. We find that super-Earth cores will take up to 30% longer to solidify than Earth's core, where this model predicts Earth's core will solidify in 6.2(3.4) billion years (Gyr) (*15*) (Fig. 3), supporting estimates for a young inner core (*35*). Owing to competing effects of stored energy versus surface area, the cores of planets smaller than Earth will solidify quickly, with the maximum time scale for solidification occurring in the 4- to 6-$M_\oplus$ size. Assuming the solidification time scale sets the time scale for dynamos, the results lead to the notable finding that super-Earths are likely to have a longer duration of magnetically shielded habitability than Earth.

REFERENCES AND NOTES

1. V. Dehant *et al.*, *Space Sci. Rev.* **129**, 279–300 (2007).
2. L. Elkins-Tanton, *Eos* **94**, 149–150 (2013).
3. A. M. Dziewonski, D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
4. R. Sinmyo, K. Hirose, Y. Ohishi, *Earth Planet. Sci. Lett.* **510**, 45–52 (2019).
5. O. L. Anderson, *Phys. Earth Planet. Inter.* **109**, 179–197 (1998).
6. R. J. Hemley, H. K. Mao, *Int. Geol. Rev.* **43**, 1–30 (2001).
7. J. Li *et al.*, *Geophys. Res. Lett.* **47**, e2020GL087758 (2020).
8. S. J. Turneaure, S. M. Sharma, Y. M. Gupta, *Phys. Rev. Lett.* **125**, 215702 (2020).
9. C. S. Yoo, N. C. Holmes, M. Ross, D. J. Webb, C. Pike, *Phys. Rev. Lett.* **70**, 3931–3934 (1993).
10. D. Valencia, R. J. O'Connell, D. Sasselov, *Icarus* **181**, 545–554 (2006).
11. C. Sotin, O. Grasset, A. Mocquet, *Icarus* **191**, 337–351 (2007).
12. A. Boujibar, P. Driscoll, Y. Fei, *J. Geophys. Res. Planets* **125**, e2019JE006124 (2020).
13. E. Gaidos, C. P. Conrad, M. Manga, J. Hernlund, *Astrophys. J.* **718**, 596–609 (2010).
14. R. G. Kraus *et al.*, *Phys. Rev. Lett.* **126**, 255701 (2021).
15. Materials and methods are available as supplementary materials.
16. J. R. Rygg *et al.*, *Rev. Sci. Instrum.* **91**, 043902 (2020).
17. H. Cao *et al.*, *Geophys. Res. Lett.* **41**, 4127–4134 (2014).
18. A. B. Belonoshko, R. Ahuja, B. Johansson, *Nature* **424**, 1032–1034 (2003).
19. W. Luo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9962–9964 (2010).
20. F. Xu *et al.*, *Earth Planet. Sci. Lett.* **563**, 116884 (2021).
21. J. H. Nguyen, N. C. Holmes, *Nature* **427**, 339–342 (2004).
22. M. W. Chase, *NIST-JANAF Thermochemical Tables*, Monograph No. 9 (American Institute of Physics, 1998).
23. S. Sugita, K. Kurosawa, T. Kadono, *AIP Conf. Proc.* **1426**, 895–898 (2012).
24. T. J. Ahrens, J. D. O'Keefe, *Moon* **4**, 214–249 (1972).
25. S. T. Stewart, A. Seifter, A. W. Obst, *Geophys. Res. Lett.* **35**, L23203 (2008).
26. R. G. Kraus *et al.*, *J. Geophys. Res.* **117**, E09009 (2012).
27. S. Anzellini, A. Dewaele, M. Mezouar, P. Loubeyre, G. Morard, *Science* **340**, 464–466 (2013).
28. G. Morard, J. Bouchet, D. Valencia, S. Mazevet, F. Guyot, *High Energy Density Phys.* **7**, 141–144 (2011).
29. L. Stixrude, *Philos. Trans. R. Soc. London Ser. A* **372**, 20130076 (2014).
30. L. Stixrude, *Phys. Rev. Lett.* **108**, 055505 (2012).
31. J. van Summeren, E. Gaidos, C. P. Conrad, *J. Geophys. Res. Planets* **118**, 938–951 (2013).
32. R. F. Smith *et al.*, *Nat. Astron.* **2**, 452–458 (2018).
33. P. J. Carter, S. J. Lock, S. T. Stewart, *J. Geophys. Res. Planets* **125**, e2019JE006042 (2020).
34. S. N. Raymond, E. Kokubo, A. Morbidelli, R. Morishima, K. J. Walsh, in *Protostars and Planets VI*, H. Beuther, R. S. Klessen, C. P. Dullemond, T. K. Henning, Eds. (Univ. of Arizona Press, 2014), pp. 595–618.
35. R. K. Bono, J. A. Tarduno, F. Nimmo, R. D. Cottrell, *Nat. Geosci.* **12**, 143–147 (2019).

2D MATERIALS

Crossover between strongly coupled and weakly coupled exciton superfluids

Xiaomeng Liu[1]†, J. I. A. Li[2]†, Kenji Watanabe[3], Takashi Taniguchi[4], James Hone[5], Bertrand I. Halperin[1], Philip Kim[1]*, Cory R. Dean[6]*

In fermionic systems, superconductivity and superfluidity occur through the condensation of fermion pairs. The nature of this condensate can be tuned by varying the pairing strength, which is challenging in electronic systems. We studied graphene double layers separated by an atomically thin insulator. Under applied magnetic field, electrons and holes couple across the barrier to form bound magneto-excitons whose pairing strength can be continuously tuned by varying the effective layer separation. Using temperature-dependent Coulomb drag and counterflow current measurements, we were able to tune the magneto-exciton condensate through the entire phase diagram from weak to strong coupling. Our results establish magneto-exciton condensates in graphene as a model platform to study the crossover between two bosonic quantum condensate phases in a solid-state system.

I n the presence of attractive interactions, a fermionic system can become unstable against pairing, forming composite bosons. These paired fermions then can form a low-temperature condensate phase. It has long been recognized that the nature of the fermionic condensate and its phase transition are directly governed by the strength of the pairing interaction U relative to the Fermi energy E_F (Fig. 1A) (*1–4*). Electrons in metals provide a paradigm example of the weak-coupling regime, where the pairing interaction is small relative to the Fermi energy ($U << E_F$). A low-temperature superconducting phase emerges from this weakly interacting Fermi liquid, described by the Bardeen-Cooper-Schrieffer (BCS) theory (*5*). In this regime,

[1]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [2]Department of Physics, Brown University, Providence, RI 02912, USA. [3]Research Center for Functional Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [4]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [5]Department of Mechanical Engineering, Columbia University, New York, NY 10027, USA. [6]Department of Physics, Columbia University, New York, NY 10027, USA.
*Corresponding author. Email: pkim@physics.harvard.edu (P.K.); cd2478@columbia.edu (C.R.D.)
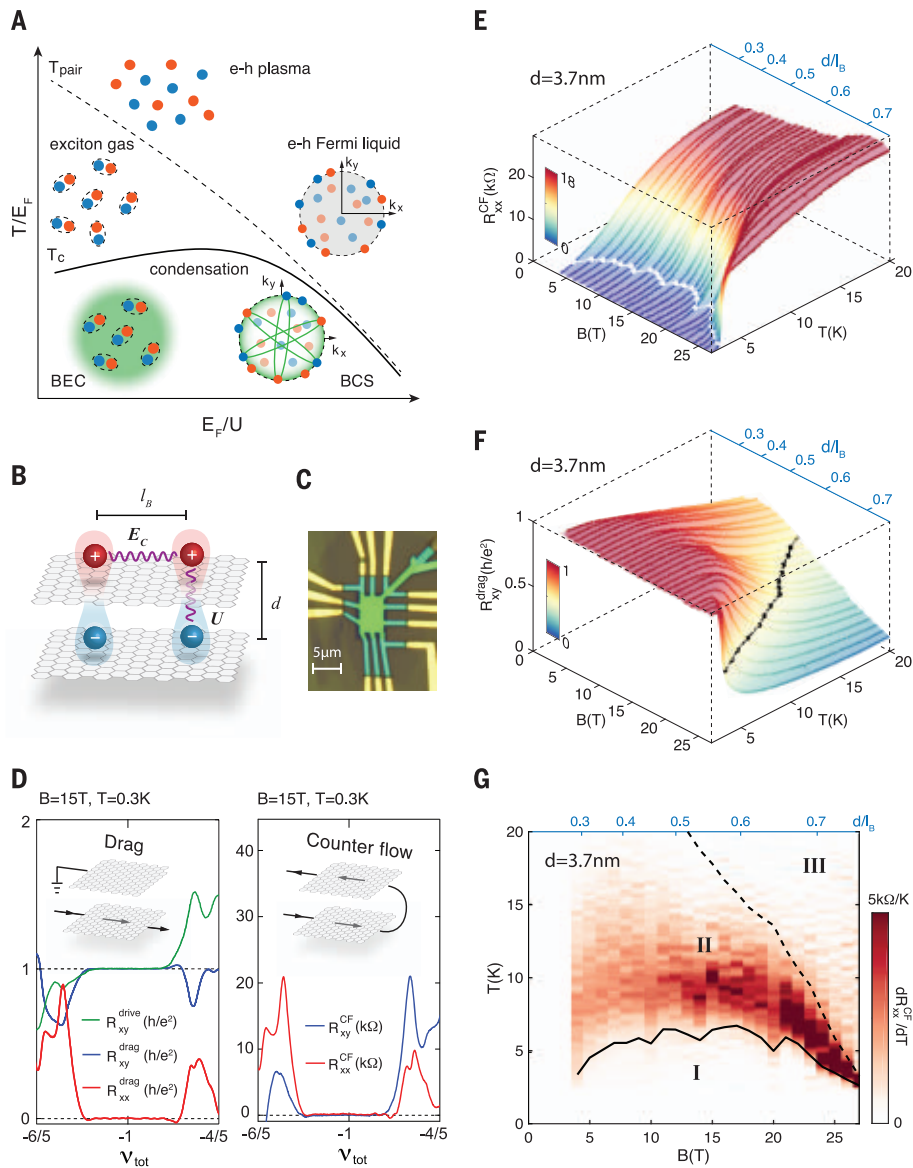†These authors contributed equally to this work.

Fig. 1. Two regimes of the exciton condensate. (**A**) Schematic phase diagram for equal densities of electrons and holes with varying temperature and coupling strength. The temperature axis is in units of the Fermi energy E_F. In the strong coupling limit ($E_F/U \ll 1$), the electrons (orange circles) and holes (blue circles) start to pair at T_{pair} and condense at much lower temperature T_c. The green halo signifies the condensate. In the weak coupling limit ($E_F/U \gg 1$), the electrons and holes exist as Fermi liquids at high temperatures and establish the BCS type of pairing below T_c. Wave vectors k_x and k_y are indicated; the green lines denote pairing between electrons and holes on the Fermi surface. (**B**) Illustration of the energy and length scales associated with exciton pairing in a graphene double-layer structure under a magnetic field. Interlayer Coulomb coupling U depends on the interlayer separation d, whereas intralayer Coulomb repulsion E_c is determined by the magnetic length l_B. (**C**) Optical image of a graphene double-layer device used in this study. (**D**) Left: Coulomb drag response of exciton condensate at $\nu_{tot} = -1$. Inset: Schematic for the drag measurement setup; arrow indicates the direction of current flow in the drive layer. Right: Longitudinal and Hall resistance in counterflow geometry measured at $\nu_{tot} = -1$. Inset: Schematic for counterflow measurement setup. Arrows indicate the direction of current flow in each layer. (**E**) Waterfall plot of longitudinal resistance from counterflow measurement as a function of temperature measured for a range of B values. The white line marks the superfluid transition temperature, T_c, where R_{xx}^{CF} drops to near zero. (**F**) Waterfall plot of Hall drag response as a function of temperature measured for a range of B values. The black dashed line marks the pairing temperature, T_{pair}, where the Hall drag is half of the quantized value. (**G**) Temperature derivative of R_{xx}^{CF} as a function of temperature T and magnetic field B. The black solid and dashed lines mark T_c and T_{pair}, respectively, according to their definitions in (E) and (F). The corresponding d/l_B value is marked on the top axis. Area I corresponds to a condensate, area II to the normal states of excitons, and area III to the normal states of disassociated electrons and holes.

electrons near the Fermi surface form pairs in momentum space, with the size of the resulting Cooper pair usually much larger than interparticle distance (2). In the opposite limit of strong interactions ($U \gg E_F$), fermions form spatially tightly bound pairs, and the size of the pair is much smaller than the average interparticle separation. In this strongly coupled limit, the system behaves like a bosonic gas or liquid, instead of like a Fermi liquid, and the low-temperature ground state is characterized by a Bose-Einstein condensate (BEC).

A crossover between the BEC and BCS regimes can theoretically be realized by tuning the ratio of U/E_F (6–8), which also corresponds to tuning the ratio of the "size" of the fermion pairs versus the interbosonic particle spacing. In solid-state systems, where the most prominent fermionic condensates (i.e., superconductors) are found, the BEC-BCS crossover paradigm is highly relevant: Whereas most metallic superconductors are understood to be in the BCS limit, some unconventional superconductors, such as the high-T_c cuprates (3, 9–11), and twisted bilayer graphene (12) are thought to reside near the crossover ($U \sim E_F$) between the BEC and BCS limits. In cold-fermion gases, continuous tuning between the weak-coupling and strong-coupling limits has been demonstrated, and the unitary crossover regime has been firmly established (13–18). Demonstration of this same crossover in a solid-state platform (i.e., within a single electronic superconductor) has been realized only recently because of the difficulty of continuously tuning the coupling strength (e.g., varying U at fixed E_F) or the electron density (varying E_F at fixed U) sufficiently while maintaining the condensate ground state (19–22).

We examined the crossover behavior of the condensate phase of magneto-excitons in quantum Hall bilayer (QHB) systems. Superfluidic magneto-exciton condensation was first realized in QHBs fabricated from GaAs heterostructures (23) and later from graphene double layers (24, 25). Here, electron-like and hole-like quasi-particles of partially filled Landau levels (LLs) reside in two parallel conducting layers. At integer values of the combined LL filling fraction $\nu_{tot} = \nu_{top} + \nu_{bot}$, where ν_{top} and ν_{bot} are respectively the filling fractions of the top and bottom layers, electrons in one layer and holes in the other layer can pair up, forming interlayer excitons that then condense into a superfluid state at low temperatures (23).

Unlike in metallic superconductors, the pairing between fermions in QHB systems is widely tunable. Because the kinetic energy of electrons is quenched in the LLs, the energetics of this system is determined by the competition between the intralayer Coulomb interaction $E_c = e^2/\varepsilon l_B$ (in Gaussian units), where $l_B = \sqrt{\hbar/eB}$ is the magnetic length, ε is the background dielectric constant, $\hbar$ is the reduced

Planck constant, e is electron charge, and B is magnetic field, and the attractive interlayer Coulomb interaction between an isolated electron and hole in the lowest LL, $U \approx (e^2/\varepsilon)/(d + 0.8 l_B)$, where d is the interlayer separation (Fig. 1B) (26). For an isolated layer with a partially filled LL, a Chern-Simons gauge transformation can turn its strongly interacting electrons characterized by E_c to a composite Fermi liquid with Fermi energy $E_F \propto E_c$ (27). In QHBs, the ratio U/E_c, which is solely determined by d/l_B, therefore provides a characterization of the relative pairing strength,

analogous to the dimensionless parameter U/E_F for generic fermionic systems with dispersive bands (23, 28, 29). For $d << l_B$, U is on the order of E_F, resulting in relatively tightly bound electron-hole pairs, which persist at temperatures well above the transition temperature where the Bose condensate disappears. For $d >> l_B$, the two layers are only weakly coupled, with each layer described by a composite Fermi liquid. In this limit, interaction between the two Fermi surfaces can lead to a pairing instability at low temperatures, resulting in a BCS-like condensate (26, 30–34).

Experimentally, d/l_B can be continuously varied in a single device, by varying the applied magnetic field B, or across multiple devices, by changing the interlayer distance d. This provides the opportunity to continuously tune through the complete condensate phase diagram. In our study, we fabricated QHBs from graphene double layers consisting of two parallel graphene layers separated by a dielectric tunneling barrier consisting of a few layers of hexagonal boron nitride (hBN) (Fig. 1B and fig. S1). We focus on the magneto-exciton condensate appearing at $\nu_{tot} = -1$, corresponding to
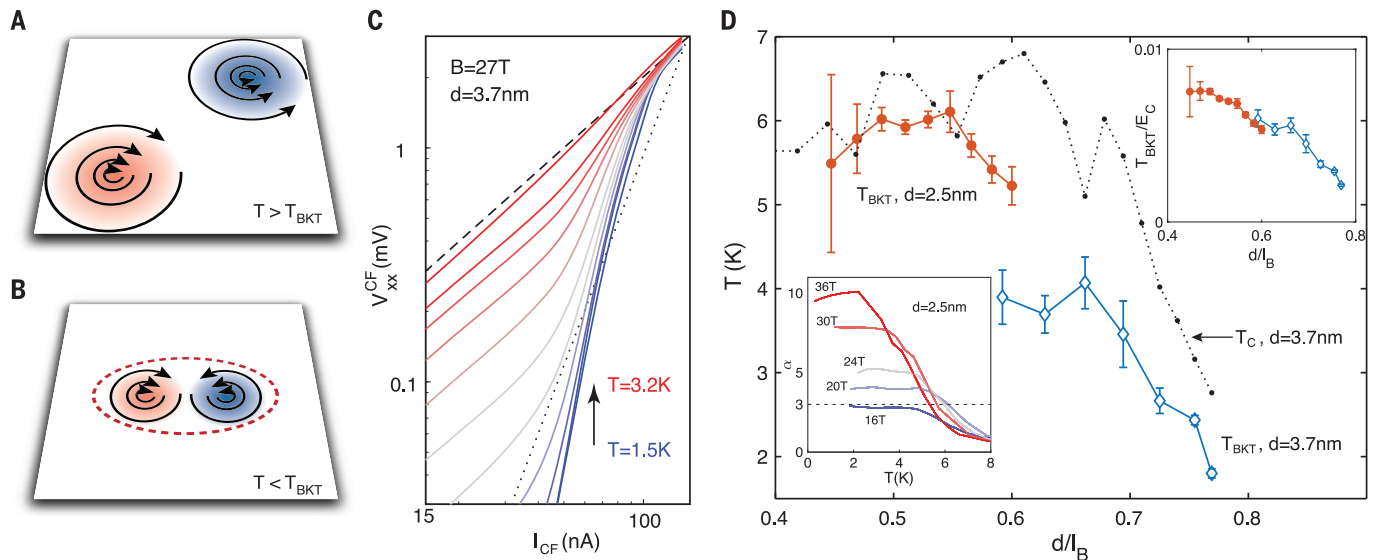
Fig. 2. BKT transition in the BCS regime. (**A** and **B**) Illustration of BKT transition. The circling black lines show the winding of the superfluid phase. Blue and red circles represent vortex and anti-vortex. When $T > T_{BKT}$, vortex and anti-vortex are free to move (A), whereas below the BKT temperature, they are bound into pairs (red dashed line) (B). (**C**) Counterflow current-voltage (I-V) relationship at $B = 27$ T in the $d = 3.7$ nm device at temperatures between $T = 1.5$ K and $T = 3.2$ K taken at approximately even temperature intervals. The dashed and dotted lines mark power-law exponents $\alpha = 1$ and 3, respectively. (**D**) BKT transition temperature as a function of d/l_B in two samples with interlayer separation of 3.7 nm and 2.5 nm (blue and red symbols, respectively). For comparison, the black dotted line shows T_c of the $d = 3.7$ nm

sample from Fig. 1E. Bottom left inset: α extracted from the I-V curves as a function of temperature for select fields in the sample with $d = 2.5$ nm. Under high magnetic fields, α rises above 3 at low temperatures, as expected for a BKT transition. However, the value of α saturates at low temperatures; as the magnetic field drops, the saturation value decreases. Eventually, for smaller magnetic fields, T_{BKT} cannot be defined, as α saturates below 3 (see, e.g., the $B = 16$ T curve). Top right inset: BKT transition temperature after scaling to Coulomb energy E_c. Data from two samples with different interlayer separation collapse onto a universal line. The error bars in the plots are estimated from the uncertainty of α obtained from power-law fitting of the I-V curves.

Fig. 3. Activation energy in the strong coupling regime.
(**A**) Arrhenius plot of R_{xx}^{CF} measured at different magnetic fields in the $d = 3.7$ nm device. (**B**) Activation gap Δ as a function of magnetic field for two devices with different interlayer separation $d = 3.7$ nm and 2.5 nm. The red solid curve corresponds to the Coulomb energy, $E_c = e^2/\varepsilon l_B$, where e is the electron charge and ε is the dielectric constant of hBN. The red dashed curve shows $0.135 E_c$.
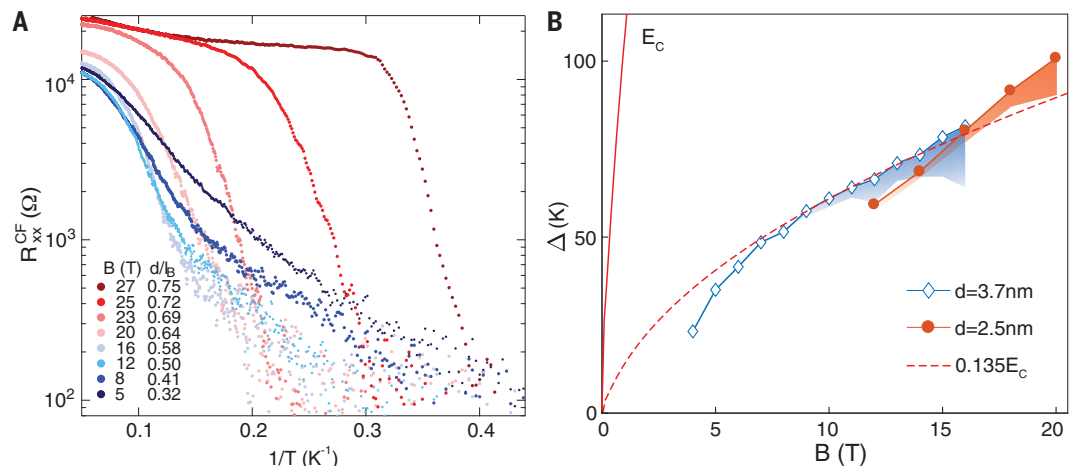
both layers filled to half filling of the first hole LL ($\nu_{\text{top}} = \nu_{\text{bot}} = -\frac{1}{2}$). We report results over the range $0.3 < d/l_B < 0.8$, where well-defined exciton superfluid states exist at the lowest experimental temperature.

To probe the dynamics of the interlayer exciton, we used the Coulomb drag and counterflow geometries (35–38) (Fig. 1D, inset) (26). In the Coulomb drag geometry, the exciton condensate is identified by the emergence of a quantized Hall resistance plateau equal to h/e^2, as measured in both the drive and drag layers, concomitant with zero longitudinal resistance in both layers (Fig. 1D). In contrast, when the two layers are decoupled, the drive layer exhibits a density-dependent Hall resistance, whereas the Hall resistance of the drag layer is close to zero (39). Thus, the Hall drag resistance R_{xy}^{drag} provides an experimental measure of interlayer pairing (23–25). In the counterflow geometry, charge-neutral excitons can be induced to flow by configuring the current to move in opposite directions in the two layers (40). In this geometry, the neutral exciton current gives a zero-valued Hall resistance in both layers, and the dissipationless nature of the superfluid condensate is revealed by a vanishing longitudinal resistance (Fig. 1D).

Figure 1, E and F, shows the temperature dependence of the counterflow longitudinal resistance R_{xx}^{CF} and the Hall drag resistance R_{xy}^{drag} of a $d = 3.7$ nm device, for a range of values of d/l_B obtained by varying the magnetic field B (see also fig. S2). At low temperatures, the exciton superfluid phase was observed over the full range of effective layer separation that we studied, $0.3 < d/l_B < 0.8$, as evidenced by the vanishing R_{xx}^{CF} in counterflow and quantized R_{xy}^{drag} (23, 36–38).

The temperature evolution of these quantities across different values of d/l_B allowed us to experimentally map key features of the condensate phase diagram. First, we identified the critical temperature of the condensate as the value below which the longitudinal transport becomes dissipationless. We defined this point as the temperature where R_{xx}^{CF} drops to less than 5% of the high-temperature saturation value. Indicated by a white line in Fig. 1E, this boundary identifies a dome below which the condensate is well formed. The dome shape of the critical temperature is consistent with theoretical expectation (29). In the strong coupling limit (small d/l_B), the primary consequence of increasing B is a corresponding increase of the exciton density ($\propto B$), which in turn drives up T_c. Conversely, in the weak coupling limit (large d/l_B), increasing d/l_B further reduces the interlayer coupling, resulting in a diminishing of the pairing between the two Fermi liquids and causing T_c to decrease.

Second, we interpret R_{xy}^{drag} as a measure of the pair fraction. In the limit of strong coupling, where electrons and holes occur in tightly bound pairs, excitons may persist at temperatures well above the counterflow-superconductivity critical temperature. In this temperature range, we would still expect to observe a large R_{xy}^{drag} response. On the other hand, at temperatures sufficiently high that electrons and holes are dissociated, the value of R^{drag} will be close to zero. We can therefore identify a temperature scale for the pair-breaking by the temperature where R_{xy}^{drag} deviates from the quantized value h/e^2. Phenomenologically, we define the pair-breaking temperature T_{pair} as the temperature where R_{xy}^{drag} drops to half its quantized value, that is, $h/2e^2$ (Fig. 1F, black line).

In Fig. 1G, we summarize the experimental phase diagram by plotting the temperature derivative of the counterflow resistance, dR_{xx}^{CF}/dT, versus d/l_B. Plotting this way emphasizes the three distinct regimes of the magneto-exciton phase diagram: the low-temperature superfluidic condensate (phase I, $T < T_c$); the intermediate phase, where there is a dissipative channel (i.e., $R_{xx}^{\text{CF}} > 0$) but the two layers remain coupled through exciton formation (phase II, $T_c < T < T_{\text{pair}}$); and the high-temperature normal phase, where the layers are decoupled and most excitons are unbound (phase III, $T > T_{\text{pair}}$). We note that the temperature range over which dR_{xx}^{CF}/dT is finite-valued tracks reasonably well the T_c and T_{pair} phase boundaries identified from Fig. 1E and Fig. 1F, respectively; this indicates that R_{xx}^{CF} and R_{xy}^{drag} are correlated in this phase diagram and that dissipation continuously increases with temperature in phase II.

The experimental phase diagram shown in Fig. 1G additionally reveals distinct temperature behavior between the small d/l_B (strong coupling) and large d/l_B (weak coupling) regimes. At small d/l_B, T_{pair} is much larger than T_c, with a gradual evolution observed between the condensate phase (phase I) and the high-temperature layer-decoupled phase (phase III). This signifies that in the strong coupling limit, the exciton pairing establishes well above the condensation temperature, consistent with the behavior expected for a BEC condensate. By contrast, at large d/l_B, T_{pair} approaches T_c, reaching toward the BCS limit. The similarity of these behaviors at small and large d/l_B to the well-known temperature dependence of the BEC and BCS limits (Fig. 1A) establishes the graphene double layer as a uniquely tunable platform where fermion pair condensation can be studied in both strong- and weak-pairing regimes (1–3, 28).

The condensate phase transitions of magneto-excitons in QHBs can be further examined in the context of two-dimensional (2D) phase transition nature. At $T < T_c$, the exciton condensate is expected to be a 2D superfluid described by the Berezinskii-Kosterlitz-Thouless (BKT) theory (41–43). To produce a counterflow voltage V_{xx}^{CF}, it is necessary that topological defects, namely vortices in the condensate order parameter (Fig. 2, A and B), should move across the sample in a direction perpendicular to the voltage gradient. Because the energy of an isolated vortex in a 2D superfluid diverges logarithmically with the size of the system, vortices can exist at low temperatures only in bound pairs of opposite signs (Fig. 2B). Counterflow resistance would not be produced by the motion of such pairs. As temperature rises, the vortices unbind at the critical temperature T_{BKT} (Fig. 2A). Above T_{BKT}, the movement of free vortices leads to a counterflow resistance. Below T_{BKT}, although the linear counterflow resistance is predicted to vanish, there can be a nonlinear response, giving a nonzero voltage at finite measuring currents. Specifically, it is predicted that for small counterflow currents I_{CF}, one should find a power-law relation: $V_{xx}^{\text{CF}} \propto (I_{\text{CF}})^\alpha$, where the exponent is given by $\alpha = 1 + [\pi\rho_s(T)/T]$, and $\rho_s(T)$ is the temperature-dependent phase-stiffness constant for the order parameter (44). According to BKT theory, $T_{\text{BKT}} = (\pi/2)\rho_s(T_{\text{BKT}})$, so α should be equal to 3 at T_{BKT} and should increase monotonically with decreasing temperature below T_{BKT} (44). In principle, the measured exponent should drop discontinuously to $\alpha = 1$ above T_{BKT}, but this decrease should be gradual for a finite measuring current.

Figure 2C plots experimental current-voltage (I-V) curves measured in the counterflow geometry in logarithmic scale. For our smallest measuring currents, below ~100 nA, we indeed observed power-law behavior, and we extracted a measured exponent $\alpha(T)$ by fitting the slope of the I-V curve at low currents. The result is plotted as a function of T in the bottom left inset of Fig. 2D. At large d/l_B, α increases with decreasing T, allowing us to extract T_{BKT} according to the criterion of $\alpha = 3$, for those α-T curves that go above $\alpha = 3$ at the lowest temperature. Figure 2D shows the experimentally obtained T_{BKT} over a large range of d/l_B for two graphene double-layer devices; in the large d/l_B limit, T_{BKT} obtained from the I-V curves follows the trend of the critical temperature T_c in Fig. 1G.

In the BCS framework, $\rho_s(T)$ collapses at the mean-field transition temperature T_m thanks to the proliferation of unpaired quasiparticles, and thus T_{BKT} is bounded by the mean-field transition temperature T_m (44). Because increasing d/l_B corresponds to weakening the interlayer BCS pairing, T_m (and thus T_{BKT}) should decline as d/l_B increases, in agreement with the experimental observation shown in Fig. 2D for $d/l_B > 0.5$. As d/l_B decreases from the BCS limit, we find that T_{BKT} first increases and then tends to saturate as the d/l_B reaches ~0.5, following the trend of T_c. Eventually the BKT transition becomes ill defined. Even for large magnetic fields, the measured value of α does not diverge as predicted for $T \to 0$, but instead saturates at a finite value (Fig. 2D,

bottom left inset). The saturation value decreases with decreasing B, and eventually falls below 3. The mechanism behind the low-temperature saturation of α is unclear but may relate to the gradual evolution of counterflow resistance as a function of temperature at small d/l_B, including possible effects of disorder. Interestingly, we find that T_{BKT} measured from two samples collapses onto a universal curve after scaling with Coulomb energy, $E_c = e^2/\varepsilon l_B$ (Fig. 2D, top right inset). This shows the critical role of Coulomb interaction in the emergence of the exciton condensate in graphene double layers.

As B decreases, we move from the BCS limit (high B) to the BEC limit (low B) and find that the transition to the low-temperature condensation phase changes qualitatively. Figure 3A shows an Arrhenius plot of R_{xx}^{CF} versus temperature at fixed values of the applied magnetic field B. Whereas at large d/l_B a sharp jump in $R_{xx}^{CF}(T)$ occurs, consistent with the BKT transition described above, at small d/l_B the counterflow resistance exhibits a thermally activated behavior $R_{xx}^{CF}(T) \sim \exp(-\Delta/2T)$ with a well-defined Δ (Fig. 3A, blue traces).

Plotting Δ as a function of B in the small d/l_B regime provides insight into the relevant low-energy excitations in the BEC limit (Fig. 3B). For both samples, the plots are well fit by $\Delta = 0.135E_c$. Qualitatively, the trend of Δ with changing B field complies with the behavior of T_{pair} shown in Fig. 1A. In the BEC limit of the illustration (Fig. 1A), T_{pair}/E_F can be approximated to a zeroth-order constant; therefore, T_{pair} is proportional to E_F. In QHB, E_c plays the role of E_F, so it is not surprising that the energy scale of pairing scales with E_c. Quantitatively, we note that this value is an order of magnitude smaller than the energy to create a free electron and hole, indicating that the appearance of the finite resistance is not caused by unbinding of excitons. The most relevant collective excitations in the small d/l_B limit are predicted to be merons and anti-merons (45), which are charged topological vortices of the exciton condensate, with large core radii (26). Merons have core energies that are a fraction of E_c, and it can be argued that in the extreme limit of $d/l_B \to 0$, there may be a regime where the density of free merons leads to $R_{xx}^{CF} \sim \exp(-\Delta/2T)$, with Δ a fraction of E_c. Our estimation of Δ for the generation of a meron–anti-meron pair is $\sim 0.6E_c$ (26); because this value is much larger than the observed Δ, disorder might play a crucial role.

We note that similar activated behavior of the counterflow current has been observed in GaAs QHBs (37, 38) in the regime of much larger d/l_B. The graphene QHB exhibits a sharp, nonactivated transition occurring in the BCS limit, where the counterflow resistance vanishes critically (fig. S2A) and the characteristic BKT type of I-V appears. These observations are absent in the GaAs QHBs. The cause of the distinct phenomenologies of the two systems remains uncertain, but we point out the following differences: The atomically thin interlayer separation of graphene QHBs allows us to access a much stronger coupling parameter range $d/l_B = 0.3$ to 0.8, as compared to $d/l_B = 1.3$ to 1.8 in GaAs (36–38, 46). The small interlayer separation in graphene QHBs makes our system less susceptible to the influence of disorder and provides activation gaps that are two orders of magnitude larger than in GaAs.

Our results show that the adjustable pairing strength in graphene double-layer structures allows access to two distinct regimes of fermion pair condensation, characterized by strong and weak coupling strength, where we uncovered distinct transport behaviors and roles of topological excitations. This dynamical and continuous tunability of fermion pairing in a solid-state device opens the door to investigating the phenomenology of fermion condensates of various pairing strengths, and may lead to improved understanding of the connection between the BCS-BEC crossover and unconventional superconductivity.

REFERENCES AND NOTES

1. A. Leggett, S. Zhang, *The BEC-BCS Crossover: Some History and Some General Observations* (Springer, 2012).
2. M. Randeria, E. Taylor, *Annu. Rev. Condens. Matter Phys.* **5**, 209–232 (2014).
3. Q. Chen, J. Stajic, S. Tan, K. Levin, *Phys. Rep.* **412**, 1–88 (2005).
4. C. A. Sá de Melo, *Phys. Today* **61**, 45–51 (2008).
5. J. Bardeen, L. N. Cooper, J. R. Schrieffer, *Phys. Rev.* **108**, 1175–1204 (1957).
6. D. M. Eagles, *Phys. Rev.* **186**, 456–463 (1969).
7. A. J. Leggett, *J. Phys. Colloq.* **41**, C7-19 (1980).
8. P. Nozières, S. Schmitt-Rink, *J. Low Temp. Phys.* **59**, 195–211 (1985).
9. M. Randeria, J. M. Duan, L. Y. Shieh, *Phys. Rev. Lett.* **62**, 981–984 (1989).
10. T. Timusk, B. Statt, *Rep. Prog. Phys.* **62**, 61 (1999).
11. J. Tallon, J. Loram, *Physica C* **349**, 53–68 (2001).
12. Y. Cao et al., *Nature* **556**, 43–50 (2018).
13. T. Bourdel et al., *Phys. Rev. Lett.* **93**, 050401 (2004).
14. C. A. Regal, M. Greiner, D. S. Jin, *Phys. Rev. Lett.* **92**, 040403 (2004).
15. M. Bartenstein et al., *Phys. Rev. Lett.* **92**, 203201 (2004).
16. M. W. Zwierlein et al., *Phys. Rev. Lett.* **92**, 120403 (2004).
17. M. G. Ries et al., *Phys. Rev. Lett.* **114**, 230401 (2015).
18. P. A. Murthy et al., *Science* **359**, 452–455 (2018).
19. S. Rinott et al., *Sci. Adv.* **3**, e1602372 (2017).
20. Y. Nakagawa et al., *Science* **372**, 190–195 (2021).
21. L. Du et al., *Nat. Commun.* **8**, 1971 (2017).
22. Z. Zhu et al., *Sci. Rep.* **7**, 1733 (2017).
23. J. Eisenstein, *Annu. Rev. Condens. Matter Phys.* **5**, 159–181 (2014).
24. X. Liu, K. Watanabe, T. Taniguchi, B. I. Halperin, P. Kim, *Nat. Phys.* **13**, 746–750 (2017).
25. J. I. Li, T. Taniguchi, K. Watanabe, J. Hone, C. R. Dean, *Nat. Phys.* **13**, 751–755 (2017).
26. See supplementary materials.
27. B. I. Halperin, P. A. Lee, N. Read, *Phys. Rev. B* **47**, 7312–7343 (1993).
28. D. Jérome, T. M. Rice, W. Kohn, *Phys. Rev.* **158**, 462–475 (1967).
29. P. B. Littlewood et al., *J. Phys. Condens. Matter* **16**, S3597–S3620 (2004).
30. N. E. Bonesteel, I. A. McDonald, C. Nayak, *Phys. Rev. Lett.* **77**, 3009–3012 (1996).
31. G. Möller, S. H. Simon, E. H. Rezayi, *Phys. Rev. Lett.* **101**, 176803 (2008).
32. G. Möller, S. H. Simon, E. H. Rezayi, *Phys. Rev. B* **79**, 125106 (2009).
33. J. Alicea, O. I. Motrunich, G. Refael, M. P. A. Fisher, *Phys. Rev. Lett.* **103**, 256403 (2009).
34. I. Sodemann, I. Kimchi, C. Wang, T. Senthil, *Phys. Rev. B* **95**, 085135 (2017).
35. J. J. Su, A. H. MacDonald, *Nat. Phys.* **4**, 799–802 (2008).
36. M. Kellogg, I. B. Spielman, J. P. Eisenstein, L. N. Pfeiffer, K. W. West, *Phys. Rev. Lett.* **88**, 126804 (2002).
37. M. Kellogg, J. P. Eisenstein, L. N. Pfeiffer, K. W. West, *Phys. Rev. Lett.* **93**, 036801 (2004).
38. E. Tutuc, M. Shayegan, D. A. Huse, *Phys. Rev. Lett.* **93**, 036802 (2004).
39. X. Liu et al., *Phys. Rev. Lett.* **119**, 056802 (2017).
40. J. P. Eisenstein, A. H. Macdonald, *Nature* **432**, 691–694 (2004).
41. V. Berezinskii, *Sov. Phys. JETP* **34**, 610–616 (1972).
42. J. M. Kosterlitz, D. J. Thouless, *J. Phys. C* **6**, 1181–1203 (1973).
43. S. M. Girvin, *Boulder School 2000: Lecture Notes* (2000); https://boulderschool.yale.edu/2000/boulder-school-2000-lecture-notes.
44. B. I. Halperin, D. R. Nelson, *J. Low Temp. Phys.* **36**, 599–616 (1979).
45. S. M. Girvin, A. H. MacDonald, *Multicomponent Quantum Hall Systems: The Sum of Their Parts and More* (Wiley-VCH, 2007).
46. T. S. Lay et al., *Phys. Rev. B* **50**, 17725–17728 (1994).
47. X. Liu et al., Replication data for "Crossover between strongly coupled and weakly coupled exciton superfluids," Harvard Dataverse (2021); doi:10.7910/DVN/4TLQYI.

ECOLOGY

The effects of defaunation on plants' capacity to track climate change

Evan C. Fricke[1,2]*, Alejandro Ordonez[3], Haldre S. Rogers[4], Jens-Christian Svenning[3]

Half of all plant species rely on animals to disperse their seeds. Seed dispersal interactions lost through defaunation and gained during novel community assembly influence whether plants can adapt to climate change through migration. We develop trait-based models to predict pairwise interactions and dispersal function for fleshy-fruited plants globally. Using interactions with introduced species as an observable proxy for interactions in future novel seed dispersal networks, we find strong potential to forecast their assembly and functioning. We conservatively estimate that mammal and bird defaunation has already reduced the capacity of plants to track climate change by 60% globally. This strong reduction in the ability of plants to adapt to climate change through range shifts shows a synergy between defaunation and climate change that undermines vegetation resilience.

Seed dispersal influences global vegetation dynamics by supporting plant regeneration and enabling species ranges to shift in response to environmental changes (1–3). Roughly half of plant species are dispersed by animals, and seed dispersal is the most widespread mutualistic function provided by vertebrates (1, 4). Therefore, disruption of the seed dispersal mutualism is a key ecological consequence of defaunation (5). The impacts of defaunation on plants are exemplified by the extinction of mutualistic interactions involving Pleistocene megafauna. The extinction of megafauna—such as the elephant-like gomphotheres in South America—not only severed coevolutionary relationships supporting plant regeneration (6, 7) but also caused large declines in long-distance dispersal, which is critical for population spread, according to dispersal reconstructions based on body mass allometries (8).

Plants today are experiencing mutualist loss as a result of past and ongoing defaunation (9) as well as climate change (10), with many populations needing to shift hundreds of meters to tens of kilometers per year to track their climatic niche (11). Further, novel communities are assembling as species introduc-

tions and shifting ranges result in cooccurrence of species that do not share coevolutionary history (12). The mutualistic interaction networks that assemble in these communities will likely influence whether certain plant species persist and spread (13). Developing the ability to predict how novel interactions and interaction extinctions affect seed dispersal function at macroecological scales is key for monitoring global human impacts on ecosystem functioning and forecasting future vegetation dynamics.

Predicting species interactions and quantifying how they affect ecosystem functioning are pressing goals for ecologists, and a growing consensus supports doing so by using a trait-based approach (14, 15). Specifically, researchers can predict interactions by modeling the links between observed interactions and species' traits such as body mass in predator-prey interactions (16) or bill and corolla length in hummingbird-plant interactions (17). Complementary data, often also related to traits, can establish how interactions translate to ecosystem functioning (18, 19). For example, traits may be used to predict whether a flower-animal interaction results in effective pollination or nectar robbing (20). The increasing

availability of data on species' traits and interactions brings global predictions within reach, and machine learning tools now offer suitable methods to capture the complex relationships that shape these interactions (17). We leverage these data and models to apply this trait-based approach to (i) make quantitative predictions of pairwise species interactions and resulting ecosystem function that can be applied at the global scale, (ii) assess the predictability of novel species interactions and their potential to constrain range shifts, (iii) estimate how altered species composition has affected seed dispersal function globally, and (iv) quantify the resulting impact on the ability of plants to track climate change.

Our first aim was to build predictive models for seed dispersal interactions involving mammals and birds, and to estimate two key functional outcomes of these interactions for plants: seed dispersal distance and the impact of gut passage on germination. By extending standard methods based on trait-matching and body mass allometries (2, 8), we developed trait-based models for discrete components of the seed dispersal process for birds and mammals (fig. S1) with boosted regression trees, a synthesis of data on the seed dispersal process, and databases of plant and animal traits (21). Analyses of the seed dispersal process involved a database of ~18,000 distinct interactions from 406 local networks where seed dispersal interactions were recorded at a given location and time period (22), data on disperser movement and seed retention time involving 302 animal species (23, 24), and the results of 2215 experiments that tested germination of seeds that were consumed and passed by frugivores versus germination of seeds untouched by

[1]National Socio-Environmental Synthesis Center, University of Maryland, Annapolis, MD 21401, USA. [2]Department of BioSciences, Rice University, Houston, TX 77005, USA. [3]Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Department of Biology, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark. [4]Department of Ecology and Evolutionary Biology, Iowa State University, Ames, IA 50011, USA.
*Corresponding author. Email: evancfricke@gmail.com

Fig. 1. Functional outcomes of plant-frugivore interactions predicted on the basis of plant and animal traits. (**A**) Mean seed dispersal distances and (**B**) germination effects (log odds ratio; values >0 indicate improvement in germination relative to baseline germination for seeds untouched by frugivores) predicted on the basis of multiple traits and plotted against jittered animal body mass. Each point represents an observed plant-frugivore interaction in the network database. [Photos left to right via Wikimedia: T. Castro, K. Schneeberger, G. Stolz, M. Thyssen, J. Boone, C. Sharp, V. Panzirsch]

frugivores (25). The cross-validated multi-trait models achieved high accuracy, strongly outperforming traditional approaches based on body mass allometries or size matching fitted by using generalized linear models. For an independent tenth of the dataset not used to fit the model, we could predict which interactions occur with 88% accuracy (table S1 and fig. S2). Further, the boosted regression tree models used to characterize seed dispersal distance and impacts of gut passage on germination also substantially out-

performed the generalized linear models (table S1). We show these functional outcomes for each interaction observed in the network database (Fig. 1), but these estimates can also be made for any other plant-animal species pair, given directly measured or phylogenetically imputed traits. Such estimates of species interactions and dispersal can be used to improve forecasting of biotic responses to climate change (26).

Having created predictive models of seed dispersal interactions on the basis of observed

interactions in existing ecosystems, we next sought to assess our ability to predict interactions in novel communities. Although the capacity to do so would advance ecological forecasting, it is currently not possible to validate predictions for future novel community scenarios. Further, the capacity to develop predictions based on current observations and standard, easily measured traits is uncertain because interaction probabilities may be determined by subtle relationships between other behavioral, chemical, and morphological factors
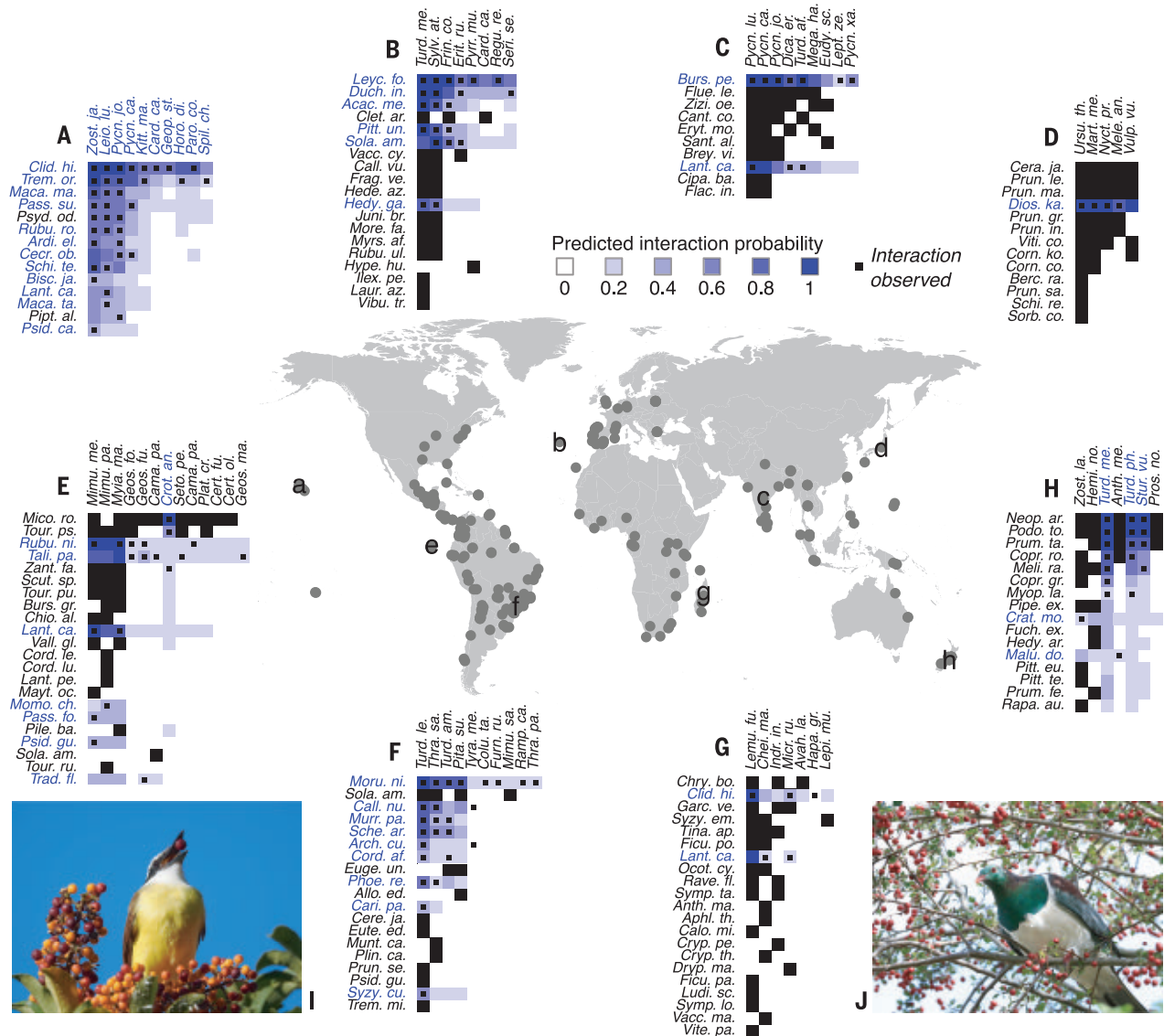
Fig. 2. Validating novel network assembly forecasts using observed interactions with recently introduced species. (**A** to **H**) Interactions between plants (rows) and animals (columns) in eight local network examples, with lowercase letters on the map showing study locations. Large black squares show observed interactions among native species. Introduced plant and animal names are shown in blue and native species in black, with the blue gradient showing the interaction probability predicted by a model trained on native species interactions only. The small black squares indicate

interactions involving introduced species observed in the field. Points show study locations for local network studies. Full species names are shown in table S2. Photos show examples of interactions involving introduced species predicted to be relatively likely and that were not observed in the local networks but have been reported elsewhere. (**I**) *Pitangus sulphuratus* and *Schefflera arboricola*, species in local network f (Photos: S. Andrade), and (**J**) shows *Hemiphaga novaeseelandiae* and *Crataegus monogyna*, species in local network h (Photos: S. Attwood).
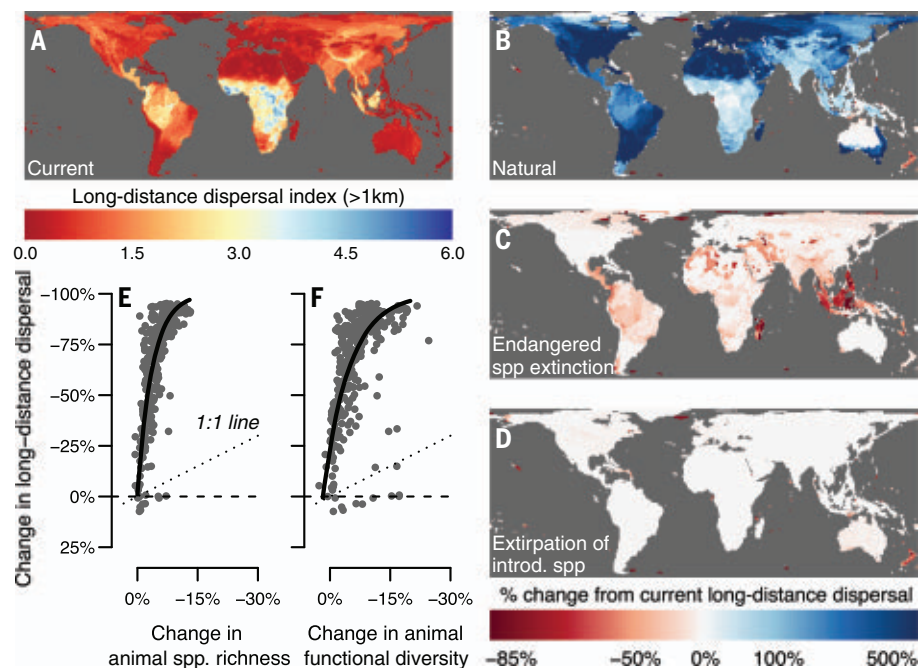
Fig. 3. Impacts of human activities on long-distance seed dispersal by birds and mammals.
(**A**) Spatial variation in the long-distance dispersal index. (**B** to **D**) Percent change in dispersal, relative to current estimates, for the natural scenario, the extinction of threatened and endangered species, and the extirpation of introduced species. (**E** and **F**) Change in long-distance dispersal function versus change in species richness and functional diversity in the current scenario relative to the natural scenario, averaged within ecoregions represented by at least 10 grid cells.

shaped by coevolution among species with long-term shared histories (27). To test our ability to predict novel interactions, we used interactions involving introduced species as a proxy for interactions in future novel communities. We trained a species interaction model using only data on interactions among species native to the local network study location, then validated this model with a test set of data involving the introduced species at each study location. Using "current" observations to predict "novel" interactions in this way (Fig. 2), we found that the accuracy of novel interaction predictions equaled or outperformed the cross-validated prediction accuracy of native species interactions (table S1), which demonstrates a strong ability to predict novel network assembly based on standard species traits. The differences that did exist between predicted and observed interactions may be explained in part by incomplete sampling of empirical networks (28). Indeed, many interactions that were predicted to occur but were not recorded in the database are known to occur from other literature and natural history observations (Fig. 2, I to J). The high capacity to predict "novel" interactions indicates that species lacking a shared phylogenetic history can interact readily and predictably using easily quantified traits. Because the same trait-matching processes that operate in native communities predict which species

interact in novel communities, more subtle factors shaped by coevolution at long temporal scales are unlikely to heavily shape network assembly—and thus limit plant range shifts—under climate change.

Our third aim was to assess how alteration of mammal and bird species composition has affected seed dispersal function globally. We focused on long-distance dispersal because of its key role in population expansion, gene flow, and recovery from landscape disturbance (2). We calculated an index of long-distance dispersal that estimates the quantity of seeds that a given assemblage of birds and mammals disperse to distances >1 km and that germinate (21). These estimates are based on models capturing the probability that each bird and mammal species participates as a seed disperser, the probability that it disperses a focal plant species, the quantity of seeds it disperses, its impact on germination through gut passage, and the distances it disperses seeds (fig. S3 and tables S1 and S3). We developed maps of long-distance dispersal under four scenarios of present-day bird and mammal species composition to assess how current long-distance dispersal function has been affected by defaunation and invasion and how it is threatened by species endangerment. First, we used current range maps that reflect the natural and introduced ranges

of extant species today (current scenario). Second, we considered a scenario representing mammal and bird ranges as they would occur today had they been unaffected by extinctions, range contractions, or introductions (natural scenario). This represents a counterfactual present-day scenario of species composition, with natural ranges constructed using occurrence records from recent history or by using fossil occurrence data and accounting for range shifts due to environmental changes (21). Third, we considered a scenario of the extinction of bird and mammal species listed as vulnerable or endangered by the International Union for Conservation of Nature Red List (endangered species extinction scenario). Fourth, we considered a scenario representing the extirpation of species from their introduced ranges (extirpation of introduced species scenario). We calculated the index of long-distance dispersal for a representative plant species with median trait values provided by the animal species present in each grid cell under each scenario.

We found that current seed dispersal function has steeply declined from its natural level (Fig. 3, A and B), with declines particularly widespread outside the tropics. The few regions where current long-distance dispersal by birds and mammals exceeds the natural level are primarily island systems with few native mammal species (Fig. 3B). In a scenario featuring the loss of vulnerable and endangered species (Fig. 3C), the greatest effect relative to current levels would occur in regions including Southeast Asia and Madagascar. This indicates that remaining animal-mediated seed dispersal function in these regions is largely provided by species that are already threatened or endangered. Extirpation of introduced species would reduce long-distance dispersal, particularly on islands (Fig. 3D). To understand how the change in dispersal function compares with the change in species richness, we calculated average values within ecoregions represented by at least 10 grid cells for the current and natural scenarios (table S4). We found that the magnitude of dispersal function loss far exceeds the magnitude of species richness loss (Fig. 3E). Ecoregions have disproportionately lost the species that provide most seed dispersal function, particularly large mammals that move over large distances (29). Further, the magnitude of function decline was also greater than that of functional diversity, measured using functional dispersion (Fig. 3F), albeit less so than when compared with richness. This shows that functional diversity reflects function better than species richness, but functional diversity can still underestimate particular functions such as long-distance seed dispersal. Large vertebrates are especially important for long-distance seed dispersal, and human activities disproportionately affect these key seed dispersers (8, 9).

Our fourth and final aim was to address the synergistic effects of defaunation and climate change by estimating how altered bird and mammal assemblages influence the capacity of plants to track local climate change through seed dispersal. Dispersal is a crucial aspect of biotic climate adaptation for plants (*30*), and dispersal limitation, along with demographic and competitive constraints, underlies lagged responses of plants to climate change (*31*). Because propagule pressure is a key factor in range expansion (*32*), we sought to quantify "climate-tracking dispersal," or the ability of seed dispersal to track climate change. We conceptualized this problem by considering how many seeds removed by frugivores from a plant during a given year's fruiting season disperse to distances that exceed the average distance that climate isoclines shift during a year—that is, climate change velocity measured in kilometers per year (*11*). For each grid cell,

we calculated the climate-tracking dispersal index based on modeled dispersal estimates involving a representative plant species with median trait values, the animal species within each grid cell, and its observed recent climate velocity. The climate change velocities within each pixel were calculated by averaging across mean annual temperature and total annual precipitation velocities observed for the period 1975–2013 (*33*).

We found that climate-tracking dispersal is especially limited in temperate regions and areas with little topographic complexity (Fig. 4A). We investigated how climate-tracking dispersal function has been affected by past defaunation and how it is threatened by the current endangerment of birds and mammals (Fig. 4B). Globally, we estimate that past species losses have caused a 59.7% average reduction in climate-tracking dispersal, and the future loss of vulnerable and endangered

species from their current ranges would result in a further reduction of 15% globally. Substantial spatial variation exists in the loss of, and threat to, climate-tracking dispersal. For example, dispersal—even by small birds—can track the relatively low climate velocities in the tropical Andes, and this region has lost little of its climate-tracking dispersal function. By contrast, climate-tracking seed dispersal is limited in regions with relatively high climate velocities, such as eastern North America and Europe, as a result of past losses of large mammals that provided long-distance dispersal. Although climate-tracking dispersal is relatively high in parts of Southeast Asia, this dispersal function is threatened by current species endangerment. Regions such as Madagascar have lost substantial climate-tracking dispersal function, and remaining function is threatened by species endangerment. Together, these results show that defaunation has already limited the ability of animal-dispersed plants in many parts of the world to keep pace with climate change and that endangered species provide much of the climate-tracking dispersal function in many other regions.

Using novel, predictive, and function-based models of animal-mediated seed dispersal, we show that publicly available data can be leveraged to accurately predict species interactions and map resulting ecosystem functions globally. This shows how shifting from a species-based to a function-based understanding of human impacts on global biodiversity can enable predictions for attributes of novel communities and direct monitoring of change in ecosystem function. Using this approach, we found that seed dispersal function globally has declined sharply from its natural level. Not only have human activities caused the climate to change rapidly—requiring broad-scale range shifts by plants—but defaunation of birds and mammals reduces the ability of plants to shift their ranges. Synergistic threats to seed disperser movement imposed by habitat fragmentation and other land-use changes (*34*) will likely amplify existing constraints on plant range shifts. This underscores the need to not only promote habitat connectivity to maximize the functional potential of current seed dispersers (*35*) but also restore biotic connectivity through the recovery of large-bodied animals (*29*) to increase the resilience of vegetation communities under climate change.
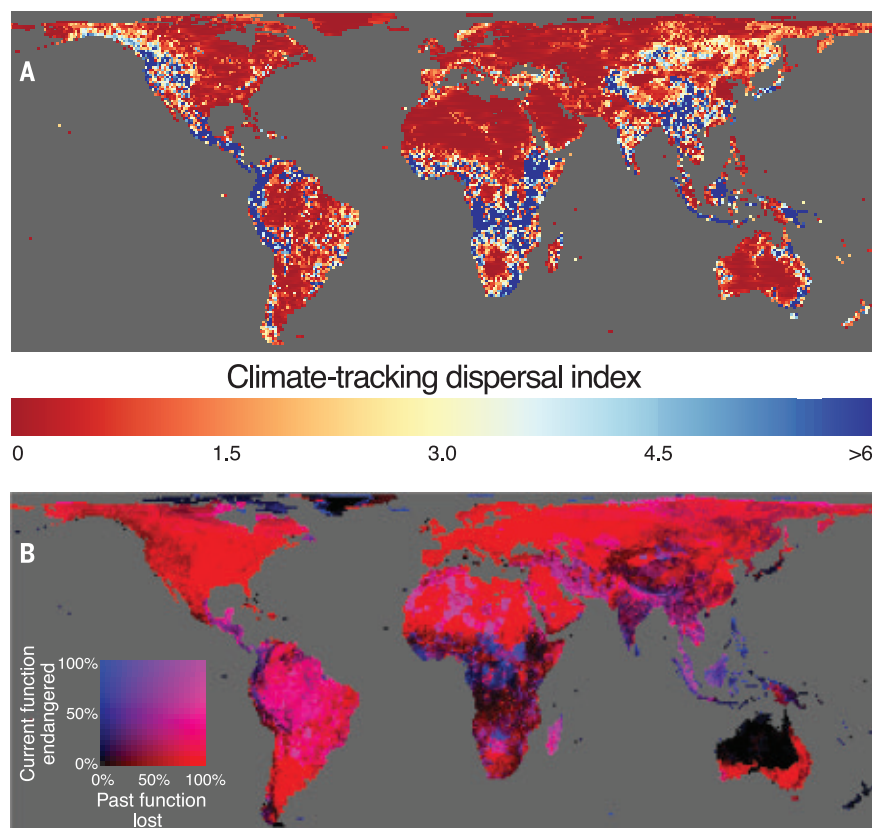
Fig. 4. Seed dispersal tracking local climate change velocities and the impacts of defaunation and species endangerment on climate-tracking dispersal. (**A**) Spatial variation in the climate-tracking dispersal index, which describes the quantity of seeds that disperse further than the local yearly climate velocity. (**B**) Loss of, and threat to, climate-tracking dispersal function caused by past defaunation and current species endangerment are shown with a bivariate color scale. Past function lost (red axis) indicates the percent decline in climate-tracking dispersal function when comparing current function to the natural level. Current function endangered (blue axis) indicates the percent of currently remaining function that would be lost if all vulnerable and endangered species were to go extinct, calculated by comparing current and endangered species extinction scenarios.

REFERENCES AND NOTES

1. P. Jordano, in *Seeds: The Ecology of Regeneration in Plant Communities* (2013), pp. 18–61.
2. F. M. Schurr et al., *Long-Distance Seed Dispersal.* L. Østergaard, Eds. (Wiley, 2018), Vol. 38, pp. 204–237.
3. J. P. González-Varo et al., *Nature* **595**, 75–79 (2021).
4. H. S. Rogers, I. Donoso, A. Traveset, E. C. Fricke, *Annu. Rev. Ecol. Evol. Syst.* **52**, 641–666 (2021).
5. K. R. McConkey et al., *Biol. Conserv.* **146**, 1–13 (2012).
6. D. H. Janzen, P. S. Martin, *Science* **215**, 19–27 (1982).
7. C. E. Doughty et al., *Ecography* **39**, 194–203 (2016).

8. M. M. Pires, P. R. Guimarães, M. Galetti, P. Jordano, *Ecography* **41**, 153–163 (2018).
9. W. J. Ripple *et al.*, *Sci. Adv.* **1**, e1400103 (2015).
10. M. Schleuning *et al.*, *Nat. Commun.* **7**, 13965 (2016).
11. S. R. Loarie *et al.*, *Nature* **462**, 1052–1055 (2009).
12. S. E. Gilman, M. C. Urban, J. Tewksbury, G. W. Gilchrist, R. D. Holt, *Trends Ecol. Evol.* **25**, 325–331 (2010).
13. K. Mokany, S. Prasad, D. A. Westcott, *Nat. Commun.* **5**, 3971 (2014).
14. I. Morales-Castilla, M. G. Matias, D. Gravel, M. B. Araújo, *Trends Ecol. Evol.* **30**, 347–356 (2015).
15. I. Bartomeus *et al.*, *Funct. Ecol.* **30**, 1894–1903 (2016).
16. U. Brose *et al.*, *Nat. Ecol. Evol.* **3**, 919–927 (2019).
17. M. Pichler, V. Boreux, A. M. Klein, M. Schleuning, F. Hartig, *Methods Ecol. Evol.* **11**, 281–293 (2020).
18. M. R. Hirt *et al.*, *Trends Ecol. Evol.* **33**, 701–712 (2018).
19. E. Harvey, I. Gounand, C. L. Ward, F. Altermatt, *J. Appl. Ecol.* **54**, 371–379 (2017).
20. S. V. Rojas-Nossa, J. M. Sánchez, L. Navarro, *Oikos* **125**, 1044–1055 (2016).
21. Methods are provided as supplementary materials.
22. E. C. Fricke, J. C. Svenning, *Nature* **585**, 74–78 (2020).
23. S. C. Chen, R. Tamme, F. J. Thomson, A. T. Moles, *Ecol. Lett.* **22**, 954–961 (2019).
24. T. Yoshikawa, K. Kawakami, T. Masaki, *Oikos* **128**, 836–844 (2019).
25. H. S. Rogers *et al.*, Effects of frugivore gut passage on seed germination: A meta-analysis and implications for conservation and future study. bioRxiv 2021.10.12.462022 [Preprint] (2021).
26. M. C. Urban *et al.*, *Science* **353**, 1113 (2016).
27. I. S. Pearse, F. Altermatt, *Ecol. Lett.* **16**, 1088–1094 (2013).
28. P. Jordano, *Funct. Ecol.* **30**, 1883–1893 (2016).
29. E. Berti, J. C. Svenning, *Glob. Ecol. Biogeogr.* **29**, 2131–2142 (2020).
30. B. Huntley, T. Webb, *J. Biogeogr.* **16**, 5 (1989).
31. J. M. Alexander *et al.*, *Glob. Change Biol.* **24**, 563–579 (2018).
32. J. L. Lockwood, P. Cassey, T. Blackburn, *Trends Ecol. Evol.* **20**, 223–228 (2005).
33. A. Ordonez, J. W. Williams, J. C. Svenning, *Nat. Clim. Chang.* **6**, 1104–1109 (2016).
34. M. A. Tucker *et al.*, *Science* **359**, 466–469 (2018).
35. J. L. McGuire, J. J. Lawler, B. H. McRae, T. A. Nuñez, D. M. Theobald, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7195–7200 (2016).
36. E. C. Fricke, evancf/global-dispersal-change: Primary release of data and code. Zenodo (2021).

ATHEROSCLEROSIS

Olfactory receptor 2 in vascular macrophages drives atherosclerosis by NLRP3-dependent IL-1 production

Marco Orecchioni[1], Kouji Kobiyama[1,2], Holger Winkels[1,3], Yanal Ghosheh[1], Sara McArdle[4], Zbigniew Mikulski[4], William B. Kiosses[4], Zhichao Fan[1,5], Lai Wen[1], Yunmin Jung[1], Payel Roy[1], Amal J. Ali[1], Yukiko Miyamoto[6], Matthew Mangan[7], Jeffrey Makings[1], Zhihao Wang[1], Angela Denn[4], Jenifer Vallejo[1], Michaela Owens[1], Christopher P. Durant[1], Simon Braumann[3], Navid Mader[8], Lin Li[9], Hiroaki Matsunami[10], Lars Eckmann[6], Eicke Latz[7], Zeneng Wang[9], Stanley L. Hazen[9,11], Klaus Ley[1,12]*

Atherosclerosis is an inflammatory disease of the artery walls and involves immune cells such as macrophages. Olfactory receptors (OLFRs) are G protein–coupled chemoreceptors that have a central role in detecting odorants and the sense of smell. We found that mouse vascular macrophages express the olfactory receptor *Olfr2* and all associated trafficking and signaling molecules. Olfr2 detects the compound octanal, which activates the NLR family pyrin domain containing 3 (NLRP3) inflammasome and induces interleukin-1β secretion in human and mouse macrophages. We found that human and mouse blood plasma contains octanal, a product of lipid peroxidation, at concentrations sufficient to activate Olfr2 and the human ortholog olfactory receptor 6A2 (OR6A2). Boosting octanal levels exacerbated atherosclerosis, whereas genetic targeting of *Olfr2* in mice significantly reduced atherosclerotic plaques. Our findings suggest that inhibiting OR6A2 may provide a promising strategy to prevent and treat atherosclerosis.

Olfactory receptors (OLFRs) (*1*) constitute the largest family of G protein–coupled receptors (GPCRs), with ~1100 genes in mice and ~400 in humans (*2*). Many OLFRs require the transporters Rtp1 and Rtp2 for surface expression (*3*). OLFRs couple through a specialized Gα subunit [Gα$_{olf}$ encoded by G protein subunit alpha L (*Gnal*)] (*4*) to adenylate cyclase 3 (*Adcy3*) (*5*, *6*). Some OLFRs are also expressed in nonolfactory tissues (*7–9*).

In atherosclerotic mice, plaques contain several phenotypically distinct macrophage subsets that behave differently (*10*, *11*). In a transcriptomic study of vascular macrophages isolated from aortas of atherosclerosis-prone *Apoe*$^{−/−}$ mice, we unexpectedly detected expression of *Olfrs*, Gα$_{olf}$ (*Gnal*), *Adcy3*, *Rtp1*, *Rtp2*, and the cyclic nucleotide-gated (CNG) ion channel subunits *Cnga1*, *Cnga2*, *Cnga3*, *Cnga4*, and *Cngb1* (*12*). Here, we focus on

Olfr2, which binds medium-chain aliphatic aldehydes such as octanal with a median effective concentration (EC$_{50}$) of ~10 µM (*13–15*). Octanal is produced by reduction of the carboxy group of octanoic acid or through lipid peroxidation during oxidative stress (*16*). We studied Olfr2 expression, function, regulation, and disease relevance in mice in vitro and in vivo and its human ortholog, OR6A2, in human macrophages.

In aortas explanted from *Apoe*$^{−/−}$ and wild-type (WT) mice, we measured *Olfr2* (fig. S1, A and B) and *Rtp1*, *Rtp2*, *Adcy3*, *Gnal*, *Cnga1*, *Cnga2*, *Cnga3*, *Cnga4*, and *Cngb1* mRNA expression by quantitative real-time polymerase chain reaction (qRT-PCR) (fig. S1A). *Olfr2* expression increased in *Apoe*$^{−/−}$ mice after 2 weeks of eating a western diet (WD), which mimics the fat and calorie content in food consumed in western countries. *Olfr2* expression did not change further at 4, 8, or 12 weeks (fig. S1B), although the canonical macrophage marker CD64 (*Fcgr1*) continued to increase (fig. S1C). Confocal microscopy and flow cytometry [fluorescence-activated cell sorting (FACS)] revealed Olfr2 protein expression in ~30% of vascular macrophages (Mφ) (Fig. 1A and figs. S2 and S3A). These data suggest that only a subset of vascular Mφ express Olfr2.

Sorted live CD45$^+$ F4/80$^+$ vascular macrophages from the aortas of *Apoe*$^{−/−}$ mice expressed *Olfr2* mRNA (fig. S3B). Other cells, such as CD45$^−$ CD31$^+$ endothelial cells (ECs) and CD45$^−$ CD31$^−$ cells including possible transdifferentiated smooth muscle cells (SMCs) (*17*, *18*), also expressed some *Olfr2* mRNA (fig. S3B). Olfr2^{GFP} mice (*19*) showed green fluorescent protein (GFP) expression in vascular Mφ (Fig. 1B and fig. S3, C to E). CD45$^−$ CD31$^+$

[1]La Jolla Institute for Immunology, La Jolla, CA 92037, USA. [2]Division of Vaccine Science, Department of Microbiology and Immunology, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. [3]Department of Internal Medicine III, Division of Cardiology, Heart Center, University Hospital of Cologne, 50937 Cologne, Germany. [4]Histology and Microscopy Core Facility, La Jolla Institute for Immunology, La Jolla, CA 92037, USA. [5]Department of Immunology, School of Medicine, UConn Health, University of Connecticut, Farmington, CT 06030, USA. [6]Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA. [7]Institute of Innate Immunity, University Hospital Bonn, 53127 Bonn, Germany. [8]Department of Cardiothoracic Surgery, Heart Center, University Hospital of Cologne, 50937 Cologne, Germany. [9]Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA. [10]Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27708, USA. [11]Heart and Vascular Institute, Cleveland Clinic, Cleveland, OH 44195, USA. [12]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA.
*Corresponding author. Email: klaus@lji.org

Fig. 1. Olfactory receptor 2 (Olfr2) is expressed in vascular macrophages (Mφ), bone marrow–derived macrophages (BMDMs), and human aortic macrophages. (**A**) $Apoe^{-/-}$ mice were fed a Western diet (WD) for 2 weeks. Confocal fluorescence micrographs are shown, depicting Olfr2 (green), CD68 (macrophage marker, magenta) immunoreactivity, and Hoechst (blue) nuclear staining in whole-mount aorta. Olfr2$^+$ CD68$^+$ cells indicated with yellow arrows, Olfr2$^+$ CD68$^-$ cells indicated with red arrows. Scale bars, 20 μm. (**B**) Olfr2^{GFP} (green), CD68 (magenta), and Hoechst (blue) in whole-mount aorta from Olfr2^{GFP} and WT mice. Scale bars, 5 μm. (**C**) OR6A2 expression measured by Affymetrix gene array (BiKE database, GSE21545) expressed as RMA (robust multi-array average; log scale) as a function of macrophage content ratio (low, <0.4, and high, >0.6; 50 subjects total) as determined by Cibersort. (**D**) Plaque composition extracted with Cibersort for representative OR6A2^{low} (left) and OR6A2^{high} (right) endarterectomy plaques. NKs, natural killer cells. (**E**) Human aorta from a surgical specimen, fixed and cut into transversal blocks. (Right column) Sections were stained for OR6A2 (AF647, green), CD68 (AF568, magenta), and nuclei (Hoechst, blue). (Left column) Control (no primary) for OR6A2 staining. 40× oil objective; scale bars, 5 μm. (**F** and **G**) Whole aortas were dissected from $Apoe^{-/-}$ mice fed a WD for 2 weeks or left untreated (vehicle control). Aortas were incubated with octanal (Oct, 10 μM), LPS (500 ng/ml), or both for 12 hours. (F) Aortic Olfr2 mRNA normalized to Gapdh, $2^{-\Delta\Delta CT}$ method. (G) Flow cytometry of Olfr2 expression in CD45$^+$ live TCRβ$^-$ CD19$^-$ F4/80$^+$ vascular Mφ, median fluorescence intensity (MFI), isotype control-subtracted (n = 5 mice per group). (**H** and **I**) BMDMs from $Apoe^{-/-}$ mice were left untreated (vehicle), incubated with octanal (Oct, 10 μM), LPS (500 ng/ml), or both for 12 hours. (H) Olfr2 mRNA normalized to Gapdh. (I) Flow cytometry of Olfr2 cell surface expression on live F4/80$^+$ BMDMs. (**J**) Confocal microscopy of BMDMs visualizing Olfr2 expression (AF488, green); treatments as indicated above photomicrographs. Scale bars, 20 μm. (**K**) Aortic roots of chimeric WT $Ldlr^{-/-}$ and $Olfr2^{-/-}$ $Ldlr^{-/-}$ mice stained for Olfr2 (AF555, green), CD68 (AF647, magenta), and Hoechst (blue). Scale bars, 250 μm. High magnification in red boxes (60× oil objective; scale bars, 10 μm). Data are presented as mean ± SEM. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. P calculated by one-way analysis of variance (ANOVA) test, Tukey's multiple comparisons test for (F) to (I), or unpaired t test with Welch correction for (C).
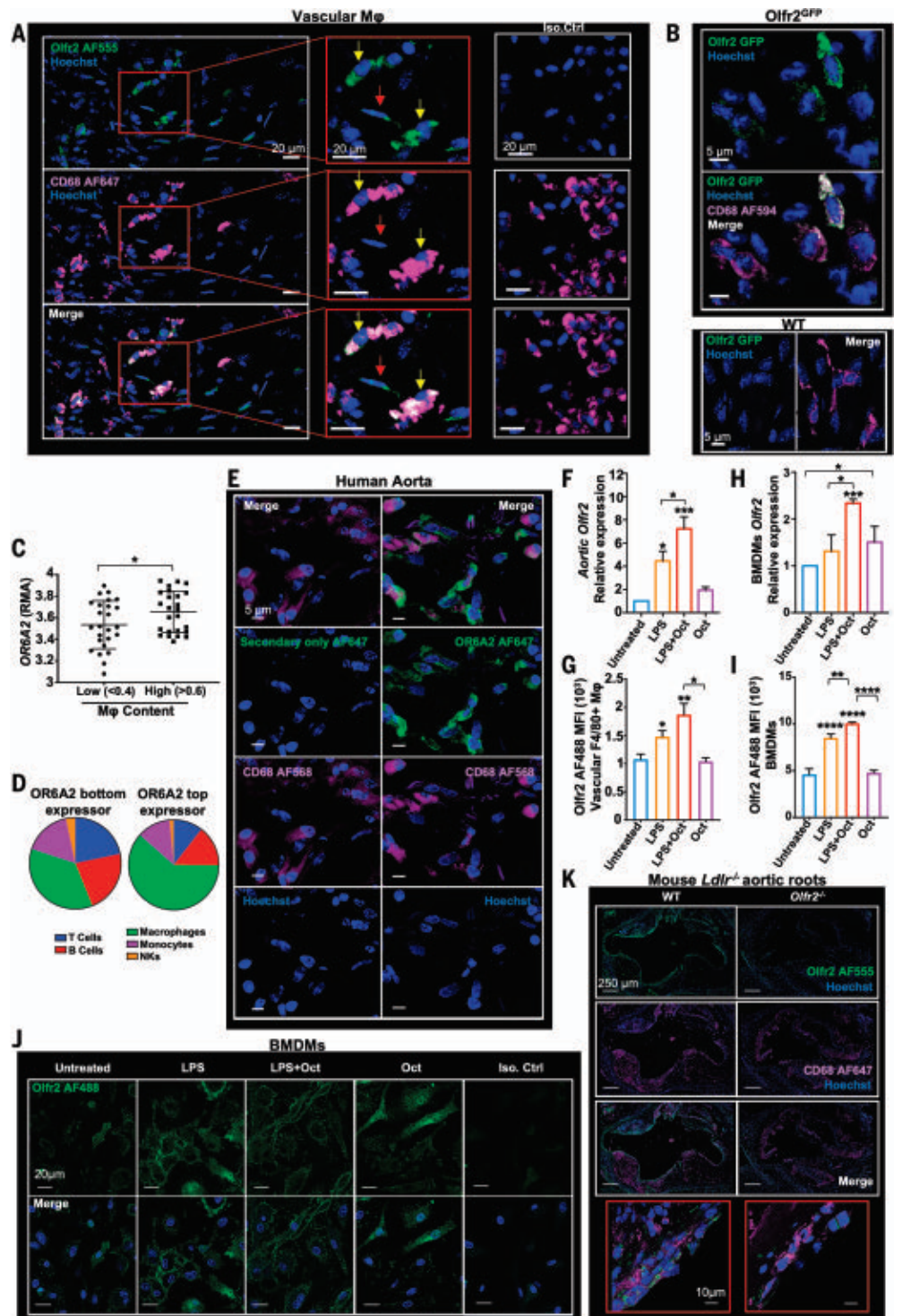
ECs also expressed some GFP, but at a lower level than macrophages (fig. S3, C to E). Antibody-based coexpression analysis showed that 80% (n = 177) of the identified cells expressing Olfr2 were CD68$^+$ macrophages (fig. S3F), with the remaining 20% of cells (n = 43) being Olfr2$^+$ CD68$^-$. Other leukocytes in the aorta showed no detectable Olfr2 expression (fig. S4).

To address expression of OR6A2 (the human ortholog of mouse Olfr2), we analyzed the BiKE (Biobank of Karolinska Endarterectomy) dataset of human carotid endarterectomy
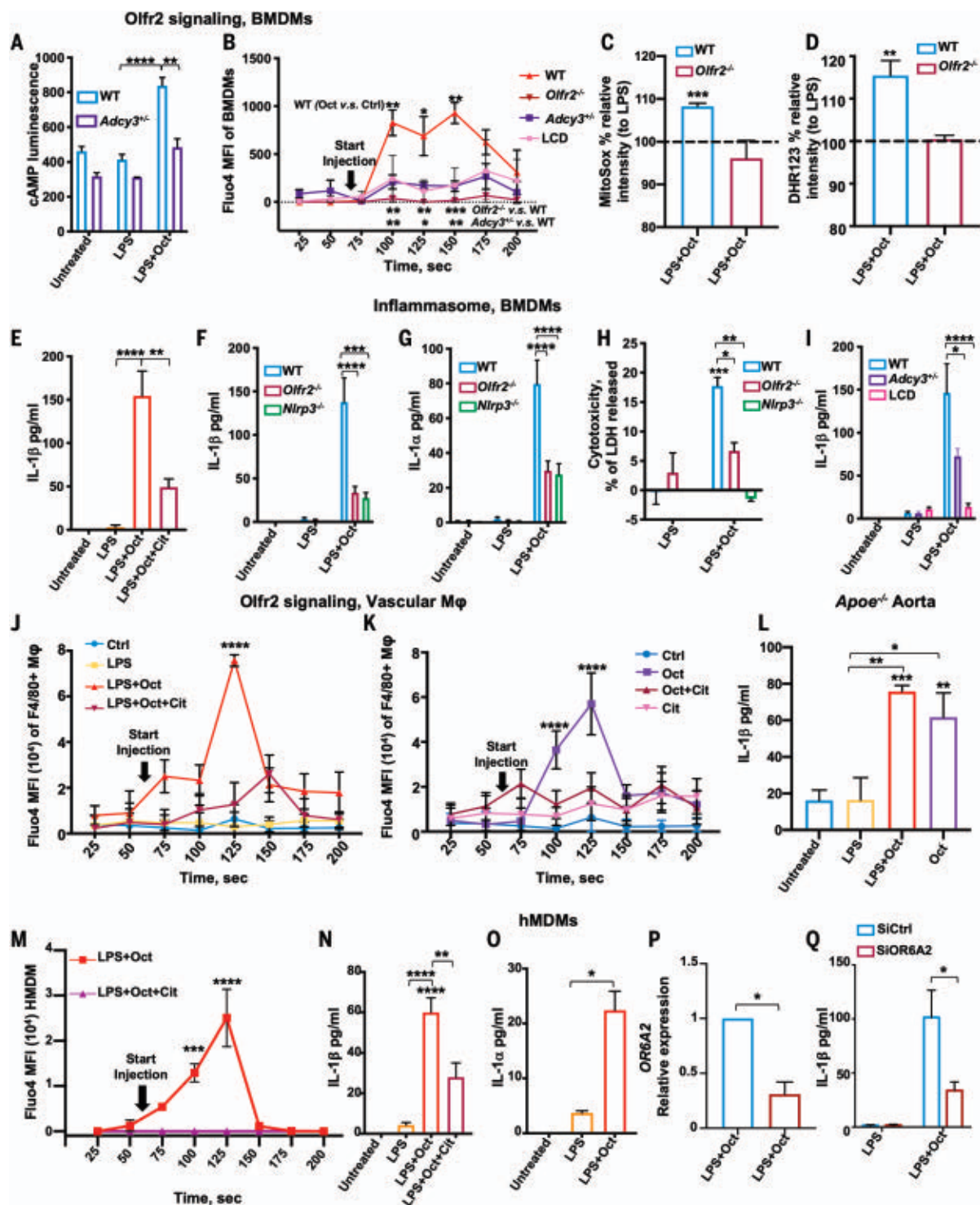
Fig. 2. Olfr2 ligation induces cAMP, Ca²⁺ flux, ROS, inflammasome activation, and IL-1 secretion in macrophages. (**A**) cAMP in BMDMs from WT and *Adcy3*⁺/⁻ mice assessed by cAMP glow assay. (**B**) Ca²⁺ flux in BMDMs from WT, *Olfr2*⁻/⁻, and *Adcy3*⁺/⁻ mice or WT pretreated for 1 hour with ʟ-*cis*-diltiazem (LCD, 100 μM). All cells were loaded with 2 μM Fluo-4, pretreated with LPS (100 ng/ml) for 1 hour, and then treated with octanal (10 μM) at "start injection." Fluo-4 MFI averaged over 25-s intervals. Three biological replicates for each time point. (**C** and **D**) BMDMs from WT or *Olfr2*⁻/⁻ mice were primed with LPS for 4 hours, treated with octanal (10 μM) for 8 hours, and then incubated with (C) 5 μM MitoSox for 30 min at 37°C or (D) 10 μM dihydrorhodamine 123 (DHR123) for 50 min at 37°C.

ROS expressed as percent of response to LPS only. (**E**) BMDMs from WT mice treated with LPS (50 ng/ml) for 4 hours followed by octanal (Oct) or octanal and citral (Oct+Cit) for 8 hours. (**F** to **H**) BMDMs from WT, *Olfr2*⁻/⁻, *and Nlrp3*⁻/⁻ mice treated with LPS (50 ng/ml) for 4 hours followed by octanal (Oct) for 8 hours. (F) IL-1β and (G) IL-1α protein in the supernatant by cytokine bead array. (H) Cytotoxicity by LDH release. (**I**) IL-1β secretion by BMDMs from WT mice, with or without pretreatment with 100 μM LCD for 1 hour, or *Adcy3*⁺/⁻ mice. (**J** and **K**) Ca²⁺ in vascular macrophages. Freshly prepared mouse aortic cell suspensions were loaded with 2 μM Fluo-4, gated for CD45⁺ live dump channel–negative (TCRb⁻, CD19⁻) F4/80⁺ and analyzed by flow cytometry. Aortic macrophages were

stimulated with LPS (100 ng/ml) for 1 hour (J) or not (K) and then treated with octanal (Oct, 10 μM), citral (Cit, 100 μM), neither (control, Ctrl), or both (Oct+Cit). (I) Mouse aortic cell suspensions from WT or $Olfr2^{-/-}$, LPS prestimulated for 1 hour and treated with octanal (Oct, 10 μM) at 60 s after the start of acquisition. Fluo-4 MFI; SD calculated for 20 to 50 cells at each time point. (L) Whole aortas were untreated (vehicle) or incubated with octanal (10 μM), LPS (500 ng/ml), or both for 12 hours (n = 5 or 6 mice per group). IL-1β protein in supernatants of stimulated aortas by cytokine bead array. (M) hMDMs were loaded with 2 μM Fluo-4, pretreated with LPS for 1 hour, and then treated with octanal (10 μM) alone or combined with citral (100 μM, Cit) at "start injection." Fluo-4 MFI averaged over 25-s intervals. Three biological

replicates for each time point. (N) IL-1β and (O) IL-1α protein in supernatants of hMDM treated with LPS (50 ng/ml) for 4 hours, left untreated, or further treated, as indicated. (P and Q) hMDMs were transfected with $OR6A2$ siRNA or scrambled control siRNAs (SiCtrl) and treated with LPS+Oct for 12 hours. (P) $OR6A2$ mRNA normalized to $GAPDH$. (Q) IL-1β protein in silenced or control hMDM treated with LPS (10 ng/ml) for 4 hours and stimulated with octanal for 8 hours. Mean ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. P calculated by two-way ANOVA test, Tukey's multiple comparisons for (A), (F) to (I), and (Q); one-way ANOVA test, Tukey's multiple comparisons test for (C) to (E), (L), (N), and (O); unpaired t test with Welch correction for (B), (J), (K), (M), and (P).

samples (GSE21545) (20) showing expression of $OR6A2$, $GNAL$, $CNGA1$, $CNGA3$, $CNGA4$, $CNGB1$, $RTP1$, and $ADCY3$ mRNA (fig. S1D). $OR6A2$ mRNA expression increased with plaque macrophage content (Fig. 1, C and D). OR6A2 protein was detectable in human aorta and colocalized with the macrophage marker CD68 (Fig. 1E). Human monocyte-derived macrophages (hMDMs) also expressed mRNA for $OR6A2$ and all signaling components (fig. S1E). $OR6A2$ mRNA was strongly increased by lipopolysaccharide plus octanal (LPS+octanal; 10 μM) (fig. S5A). OR6A2 protein was detectable by FACS (fig. S5B) and confocal microscopy (fig. S5C).

Toll-like receptor 4 (TLR4) ligands have a known role in atherosclerosis (21). We observed $Olfr2$, $Rtp1$, and $Rtp2$ expression in vascular Mφ to be significantly increased after treatment with the TLR4 agonist LPS and further enhanced in the presence of octanal (10 μM) (Fig. 1, F and G, and fig. S1, F and G). Mouse bone marrow–derived macrophages (BMDMs) showed similar patterns of expression (Fig. 1, H to J, and fig. S1H). To test the impact of Olfr2 on atherosclerosis, we generated $Olfr2^{-/-}$ mice using CRISPR-Cas9 with a guide RNA targeting exon 1 (fig. S6A), resulting in complete absence of $Olfr2$ mRNA (fig. S6B). We confirmed that Olfr2 protein in BMDMs (fig. S6C), vascular Mφ (fig. S3A), and aortic root sections (Fig. 1K) was sharply reduced in $Olfr2^{-/-}$ mice. To test the functionality of Olfr2, we studied its proximal signaling cascade (6). In mouse BMDMs, LPS+octanal, but not LPS alone, increased cyclic adenosine monophosphate (cAMP), which was sharply reduced in $Adcy3^{+/-}$ cells (Fig. 2A). $Adcy3$ encodes the adenylate cyclase responsible for production of cAMP, which activates CNG channels. Ca^{2+} flux in BMDMs in response to octanal was sharply reduced in $Olfr2^{-/-}$ and $Adcy3^{+/-}$ BMDMs and by the CNG channel inhibitor L-cis-diltiazem (LCD) (22) (Fig. 2B). Taken together, these data show that Olfr2 is functional in mouse macrophages.

RNA sequencing analysis of octanal-treated BMDMs showed that Olfr2 can activate oxidative stress pathways (fig. S7). Octanal triggered Olfr2-mediated production of mitochondrial and cytosolic reactive oxygen species (ROS)

(Fig. 2, C and D). ROS can serve as signal 2 to trigger NLR family pyrin domain containing 3 (NLRP3) inflammasome formation and activation (23). LPS served as signal 1 (gene expression of inflammasome components) (fig. S7, A to C). NLRP3-dependent caspase-1 directly cleaves and activates interleukin (IL)–1β (24) and indirectly [through calpain and gasdermin D (GSDMD)] allows release of IL-1α (25). We tested the role of Olfr2 in NLRP3 inflammasome assembly and release of IL-1β, IL-1α, and lactate dehydrogenase (LDH). The release of IL-1β and IL-1α protein increased with octanal dose in LPS-primed BMDMs (fig. S8, A and B) and was inhibited by the Olfr2 inhibitor citral (13) (Fig. 2E and fig. S8D). Knocking out $Olfr2$ or $Nlrp3$ (Fig. 2, F and G) significantly reduced IL-1β and IL-1α release by ~80%. Knocking out $Rtp1$ and $Rtp2$ (fig. S8F) or blocking the Nlrp3 inflammasome with the small-molecule inhibitor MCC950 (26), or caspase 1 with VX-765 (27), or GSDMD with ouabain (28) (fig. S8, G to I) significantly reduced IL-1β secretion. LDH release was reduced in $Olfr2^{-/-}$ and $Nlrp3^{-/-}$ BMDMs (Fig. 2H). Blocking the CNG calcium channels by LCD (Fig. 2I) or using the calcium chelator BAPTA-AM (fig. S8J) blocked IL-1β release. In $Adcy3^{+/-}$ BMDMs, IL-1β release was also significantly reduced (Fig. 2I). Octanal promoted calcium flux in vascular macrophages in freshly explanted $Apoe^{-/-}$ aortas with (Fig. 2J) or without (Fig. 2K) LPS, but not in other leukocytes and CD45^{-} cells (fig. S9, A to C). Vascular macrophages from $Ldlr^{-/-}$ mice receiving $Olfr2^{-/-}$ bone marrow and fed a high-cholesterol diet (HCD) for 8 weeks showed no calcium flux after octanal (fig. S10A). HCD is used to induce atherosclerosis in $Ldlr^{-/-}$ mice. Atherosclerotic $Apoe^{-/-}$ aortas showed increased release of IL-1β in response to octanal alone without LPS (Fig. 2L), showing that endogenous TLR ligands (21) are sufficient to allow octanal to induce IL-1β secretion. LPS treatment of atherosclerotic aortas boosted the octanal-mediated induction of other proinflammatory cytokines (fig. S10, B to I).
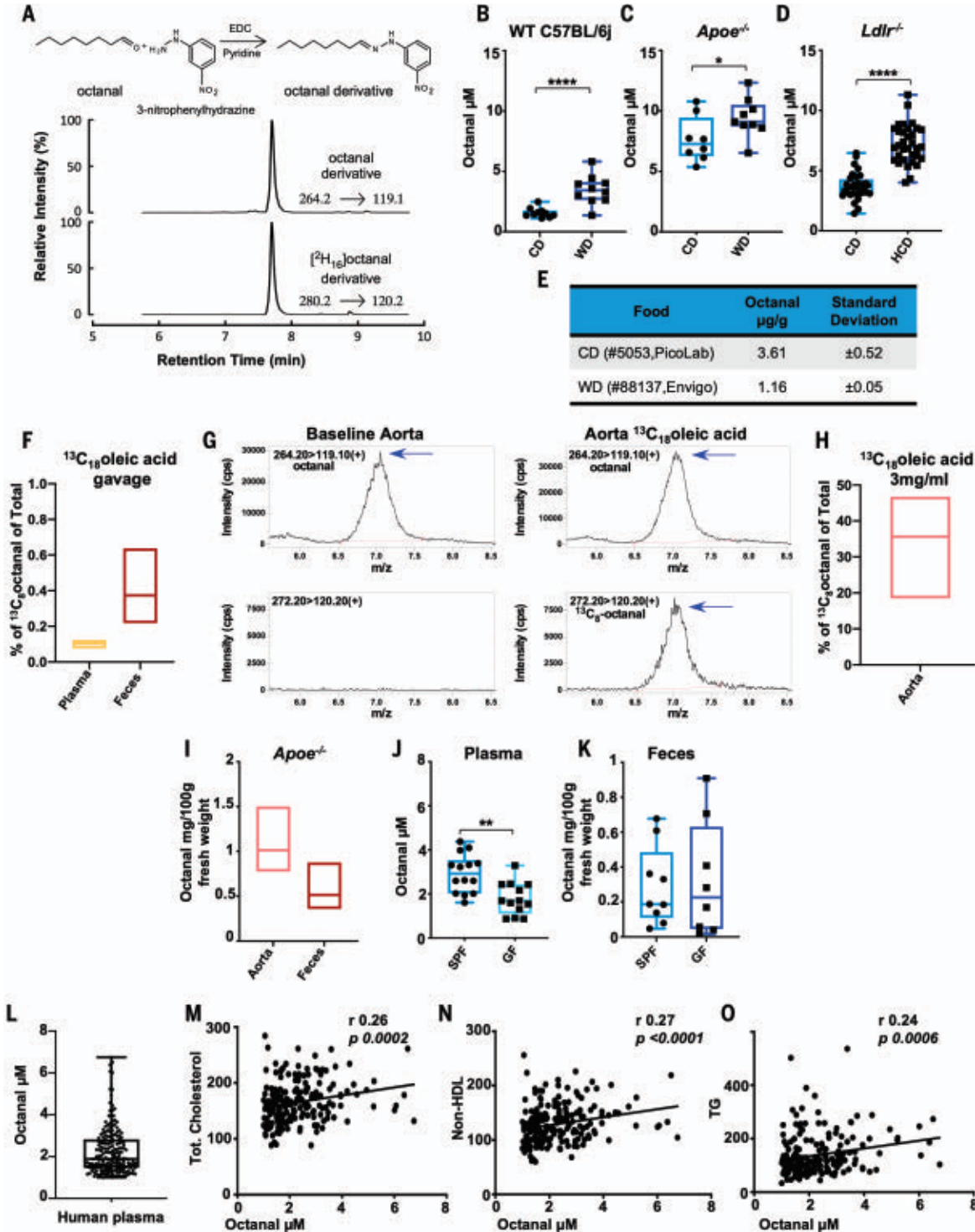
To translate these findings to human macrophages, we measured Ca^{2+} flux in hMDMs in response to LPS+octanal (Fig. 2M) and found Ca^{2+} flux to be completely abolished by citral

(Fig. 2M). hMDMs secreted IL-1β (Fig. 2N) and IL-1α (Fig. 2O) in response to LPS+octanal. Knocking down $OR6A2$ by small interfering RNA (siRNA) significantly reduced $OR6A2$ mRNA expression (Fig. 2P) and IL-1β secretion in response to octanal (Fig. 2Q). Inhibition of caspase 1 (27) or GSDMD (28) (fig. S11, A and B) significantly reduced IL-1β secretion. TNF and $IL6$ mRNA and protein were also induced by LPS+octanal (fig. S11, C to H). Thus, upstream Olfr2 signaling in mouse BMDMs and vascular macrophages is similar to olfactory receptor signaling in olfactory epithelium (6), but Olfr2 ligation in mouse or OR6A2 ligation in human macrophages culminates in NLRP3-dependent IL-1α and IL-1β release.

Octanal is known to be produced from lipid peroxidation (16) and has been detected in oxidized low-density lipoprotein (oxLDL) (29). The octanal present in oxLDL activated Ca^{2+} flux in LPS-primed WT, but not $Olfr2^{-/-}$ BMDMs (fig. S9, E and F). To measure octanal in body fluids, tissues, and feces, we developed a 3-nitrophenylhydrazine derivatization (Fig. 3A) allowing detection of octanal by mass spectrometry. Plasma from WT C57BL/6 mice eating a chow diet (CD) contained just under 2 μM octanal, which was doubled by feeding a high-fat western diet (WD, Fig. 3B). $Apoe^{-/-}$ mouse plasma contained ~7 μM octanal, which was further increased to ~9 μM by feeding a WD (Fig. 3C). Similar results were obtained in $Ldlr^{-/-}$ mice (Fig. 3D). However, the WD does not contain more octanal than the CD (Fig. 3E), implying that the relevant octanal source was not directly from food.

Lipid peroxidation of oleic acid can be a source of octanal (30, 31). To test this, we gavaged mice with 3 mg $^{13}C_{18}$ oleic acid. After 24 hours, a small fraction of octanal in plasma (<0.1%) and feces (<1%) was $^{13}C_8$ octanal (Fig. 3F and fig. S12, A to E), consistent with most oleic acid being endogenously produced in enterocytes by stearoyl-CoA desaturase-1 (SCD1) (32). Lipid peroxidation is known to occur in the atherosclerotic aorta (33). To determine whether octanal can be produced in situ in the vasculature, we cultured atherosclerotic $Apoe^{-/-}$ aortas in the presence of 3 mg/ml of $^{13}C_{18}$ oleic acid for 12 hours. High levels (~35%) of $^{13}C_8$ octanal in aorta were detected (Fig. 3,

Fig. 3. Octanal is present in mouse and human plasma and increases with high-fat diet. (**A**) Murine and human blood plasma was analyzed for octanal by derivatization and stable isotope dilution liquid chromatography with tandem mass spectrometry. [^{2}H$_{16}$] octanal was spiked into plasma before the derivatization reaction as internal standard. (**B**) Male WT C57BL/6J, (**C**) $Apoe^{-/-}$, and (**D**) $Ldlr^{-/-}$ mice were fed a chow diet (CD), western diet (WD), or high-cholesterol diet (HCD). Plasma was analyzed for octanal concentrations [n = 9 or 10 mice per group for (B) and (C), and n = 19 or 20 mice for (D)]. (**E**) Octanal concentration in pellets of CD and WD mouse food (n = 3 pellets). (**F**) Percentage of ^{13}C$_8$ octanal versus total octanal in blood plasma and feces of mice (n = 5 mice) gavaged with 3 mg ^{13}C$_{18}$ oleic acid. (**G** and **H**) ^{13}C$_8$ octanal detection in aorta harvested from $Apoe^{-/-}$ mice (n = 4 mice) and incubated with or without (baseline) 3 mg/ml ^{13}C$_8$ oleic acid for 12 hours. (G) Extracted chromatograms in positive-ion multiple reaction monitoring (MRM) mode of octanal (left) and ^{13}C$_8$ octanal (right) after reaction with 3-nitrophenylhydrazine with parent to daughter transitions, 264.2→119.1, 272.2→120.2, respectively. (H) ^{13}C$_8$ octanal, percent of total octanal detected in $Apoe^{-/-}$ mouse aortas. (I) Octanal in snap frozen aorta and feces from $Apoe^{-/-}$ mice on a CD (n = 3 mice each). Octanal in plasma (J) and feces (K) of germ-free (GF, n = 14) and conventional C57BL/6J (SPF, n = 14) mice. (L) Human blood plasma analyzed for octanal concentrations (n = 196 human specimens). Correlation of plasma octanal with (**M**) total cholesterol, (**N**) non-HDL cholesterol, and (**O**) triglycerides (TG). P values and correlation coefficients (r) were calculated by Spearman's rank analysis. Data are presented as mean ± SEM, unless otherwise specified. *P < 0.05, **P < 0.01, ****P < 0.0001. P calculated by unpaired t test with Welch correction.

G and H), suggesting that octanal can derive from oleic acid in the atherosclerotic aorta. However, whether oxLDL is a major source of octanal is still not entirely clear.

Octanal was detected in atherosclerotic $Apoe^{-/-}$ aorta at an average concentration of 1 mg/100 g tissue fresh weight (Fig. 3I), comparable to ~75 μM. We also found octanal in feces of

atherosclerotic $Apoe^{-/-}$ mice at a concentration of 0.5 mg/100 g fresh weight (Fig. 3I), comparable to ~18 μM. Thus, there is much more octanal in the aorta and feces than in

Fig. 4. Octanal supplementation exacerbates atherosclerosis, whereas genetic targeting of *Olfr2* ameliorates atherosclerosis. (**A**) Treatment protocol for (B) to (E). (**B**) Octanal in the plasma before treatment (baseline) and after 4 weeks of octanal treatment (*n* = 14 mice per group). (**C**) Pinned aortic arches of vehicle- or octanal-treated mice. Scale bars, 5 mm. (**D**) En face atherosclerotic lesion size in aortic arches, percent of area (*n* = 10 or 11 mice per group). (**E**) Plaque area (square micrometers) in aortic root serial sections from the same mice. (**F**) Treatment protocol for (G) to (O). *Ldlr*[−/−] mice (*n* = 19 or 20 mice per group) were lethally irradiated and reconstituted with WT or *Olfr2*[−/−] bone marrow and fed a HCD for 12 weeks. (**G**) Pinned aortic arches of chimeric WT *Ldlr*[−/−] and *Olfr2*[−/−] *Ldlr*[−/−] mice. Scale bars, 5 mm. (**H**) Aortic arch lesion size as percent of analyzed arch area. (**I**) Aortic root sections starting from the valve plane stained for Oil Red O (ORO), representative out of 100 sections analyzed per group (*n* = 19 or 20 mice per group; distance from valve plane in micrometers). Plaque area (square micrometers) for (**J**) total

sections and (**K**) each aortic root section as a function of position (*n* = 19 or 20 mice per position). Root lesions (G) to (K) (*n* = 11 or 12 mice per group): (**L**) Necrotic core area per lesion, H&E. (**M**) Total number (#) of CD68+ (magenta) macrophages per lesion. (**N**) Total number of smooth muscle cells per lesion (αSMA+ cells, yellow; Hoechst, blue). (**O**) Collagen content by picrosirius red (PSR) and circular dichroism. Scale bars, 500 μm. (**P**) Treatment protocol for (Q) to (S). *Ldlr*[−/−] mice (*n* = 11 to 13 per group) were lethally irradiated and reconstituted with WT or *Olfr2*[−/−] bone marrow. Mice were fed a HCD for 8 weeks, injected every 3 days intraperitoneally with octanal (Oct, 10 μg per
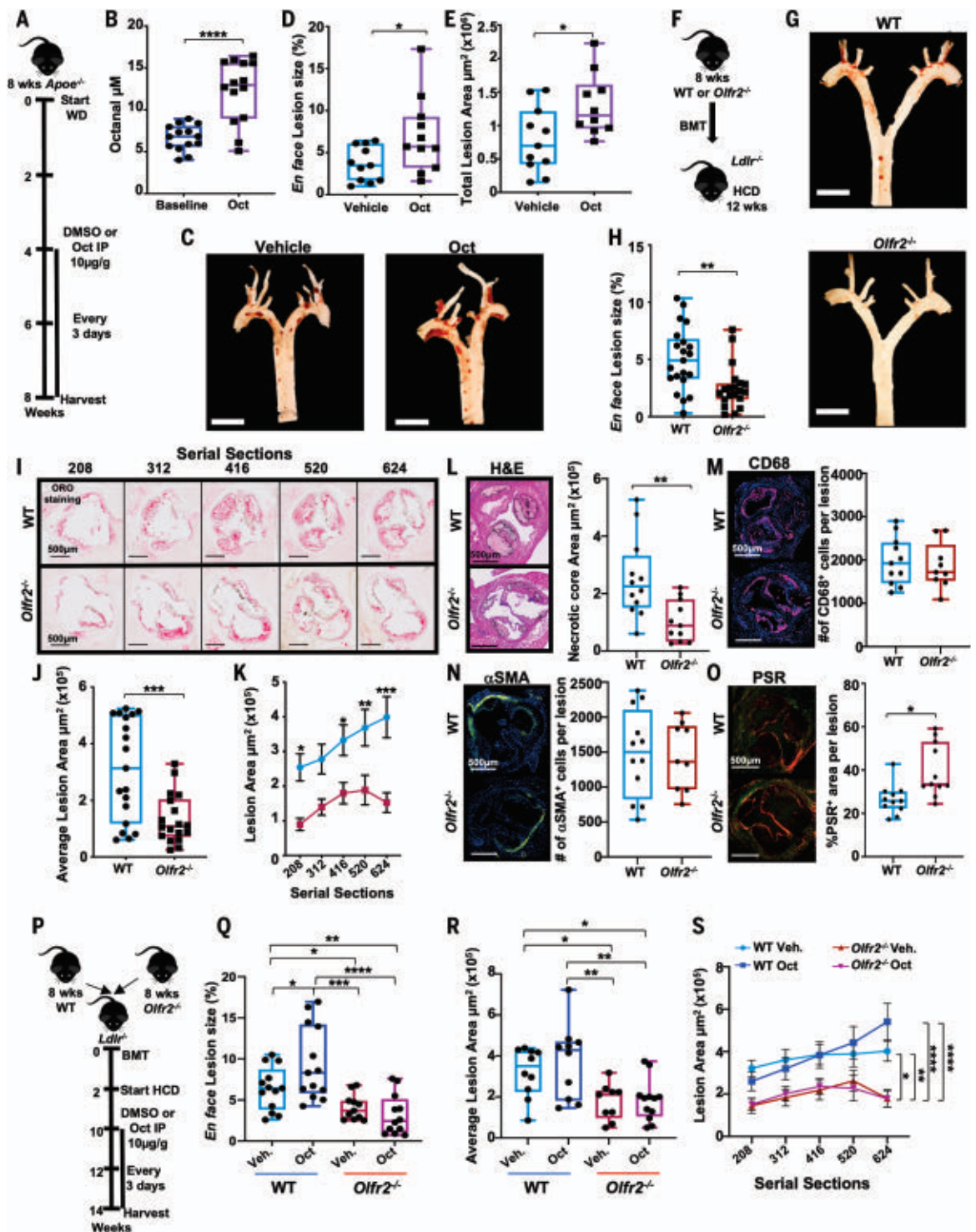
gram of body weight) or vehicle for the remaining 4 weeks of the study. (**Q**) Aortic arch lesions en face, percent of analyzed area. (**R**) Plaque area (square micrometers) quantification for root sections (*n* = 9 to 12 per group) and (**S**) each aortic root section as a function of position. Data are presented as mean ± SEM. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. *P* calculated by two-way ANOVA and Tukey's multiple comparisons test for (K) and (Q) to (S) [only the comparisons vehicle (Veh.) versus octanal (Oct.) for section 624 are reported; for the full comparison analysis, see fig. S17B] and unpaired *t* test with Welch correction for (B), (D), (E), (H), (J), and (L) to (O).

plasma. Plasma octanal was significantly reduced in germ-free (GF) mice (Fig. 3J). However, feces from GF mice contained similar levels of octanal as feces from specific pathogen–free (SPF) controls (Fig. 3K), suggesting that gut microbiota are not needed for octanal production. To translate these findings to humans, we measured octanal in 196 subjects (tables S1 and S2) and found octanal levels similar to those in mice (Fig. 3L). Octanal was significantly positively correlated with total cholesterol, non–high-density lipoprotein (HDL) cholesterol, LDL, and triglycerides (Fig. 3, M to O, and table S2). These data support the idea that octanal could be relevant for atherosclerosis in both mice and humans.

To directly test the impact of octanal in vivo, we used three mouse models: $Ldlr^{-/-}$ mice reconstituted with $Olfr2^{-/-}$ or $Rtp1/2^{dKO}$ bone marrow, in each case compared with $Ldlr^{-/-}$ mice receiving WT bone marrow, and $Apoe^{-/-}$ mice treated with octanal. $Apoe^{-/-}$ mice fed a WD for 4 weeks were injected with octanal (10 μg per gram of body weight intraperitoneally) every 3 days for the last 4 weeks of the study (Fig. 4A), which more than doubled octanal levels in the plasma (Fig. 4B). En face lesion staining of aortic arches and cross sections of aortic roots (34) showed that octanal treatment more than doubled lesion size (Fig. 4, C to E). Octanal treatment induced a systemic increase of tumor necrosis factor (TNF) and IL-1β levels in the plasma (fig. S13A) but had no effect on total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, or other blood parameters (fig. S13B). $Apoe^{-/-}$ mice treated with citral (10 μg per gram of body weight, 12 injections over 4 weeks; fig. S14A) showed ~40% reduction in atherosclerotic lesion size (fig. S14, B and C) with no effect on body weight, lipids, or leukocytes (fig. S14D). These findings suggest that endogenous levels of octanal are sufficient to induce a pathophysiologically relevant activation of Olfr2 that exacerbates atherosclerosis, which is amplified by boosting octanal levels.

To directly test the impact of Olfr2 on atherosclerosis, we reconstituted $Ldlr^{-/-}$ mice with $Olfr2^{-/-}$ bone marrow (Fig. 4F). These mice developed ~50% smaller en face aortic lesions on a HCD than littermate controls reconstituted with WT bone marrow (Fig. 4, G and H). Body weight, blood lipids, and leukocyte counts were unaffected (fig. S15). From the same mice, we prepared serial sections of aortic roots (Fig. 4I), starting from the valve plane (34). Oil Red O positively stained lesion area was significantly larger in $Ldlr^{-/-}$ mice receiving WT bone marrow than those receiving $Olfr2^{-/-}$ bone marrow, measured as total lesion area (Fig. 4J) or as a function of the distance from the valve plane (Fig. 4K). The necrotic core area detected by hematoxylin and eosin (H&E) staining was significantly reduced in $Olfr2^{-/-}$ bone marrow recipient mice (Fig. 4L), accounting for most

of the reduction in the atherosclerotic lesion size. Macrophage and smooth muscle content (Fig. 4, M and N) remained unchanged. $Olfr2^{-/-}$ lesions showed significantly increased collagen content (Fig. 4O).

$Ldlr^{-/-}$ mice lethally irradiated and reconstituted with bone marrow cells from $Rtp1/2^{dKO}$ mice (dKO, double knockout) (35) showed significantly reduced Olfr2 expression and aortic arch lesions compared with WT controls (fig. S16, A to C). Body weight, blood lipids, and leukocytes were unaffected (fig. S16D). Bone marrow transplanted $Ldlr^{-/-}$ mice fed a HCD and treated with octanal for 4 weeks (Fig. 4P) showed significantly increased lesion size en face (Fig. 4Q) in WT, but not $Olfr2^{-/-}$, recipients. Atherosclerotic lesions en face and in aortic root serial sections (Fig. 4, Q to S) were significantly smaller in $Olfr2^{-/-}$ recipients, demonstrating that Olfr2 is a disease-relevant octanal receptor in vivo. Body weight, blood lipids, and leukocyte counts were unaffected (fig. S17).

Here, we have demonstrated that the octanal receptor Olfr2 in mouse and OR6A2 in human vascular macrophages in conjunction with TLR4 ligation induces inflammasome activation in response to octanal, leading to IL-1α and IL-1β protein production and secretion. This, together with other inflammatory cytokines, likely explains the large impact of Olfr2 in mouse models of atherosclerosis. The source of octanal appears to be lipid peroxidation from oleic acid, which is most pronounced in the atherosclerotic aorta. We propose that drug-like small molecules targeting OR6A2 and possibly other OLFRs may constitute novel therapeutic targets for the treatment, prevention, and reversal of atherosclerosis.

REFERENCES AND NOTES

1. L. Buck, R. Axel, *Cell* **65**, 175–187 (1991).
2. S. Rouquier, A. Blancher, D. Giorgi, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2870–2874 (2000).
3. H. Saito, M. Kubota, R. W. Roberts, Q. Chi, H. Matsunami, *Cell* **119**, 679–691 (2004).
4. F. Li *et al.*, *J. Neurosci.* **33**, 7975–7984 (2013).
5. C. S. Silva Teixeira, N. M. Cerqueira, A. C. Silva Ferreira, *Chem. Senses* **41**, 105–121 (2016).
6. G. Antunes, F. M. Simoes de Souza, *Methods Cell Biol.* **132**, 127–145 (2016).
7. I. Ferrer *et al.*, *Front. Aging Neurosci.* **8**, 163 (2016).
8. S. J. Lee, I. Depoortere, H. Hatt, *Nat. Rev. Drug Discov.* **18**, 116–138 (2019).
9. J. J. Li *et al.*, *PLOS ONE* **8**, e80148 (2013).
10. C. Cochain *et al.*, *Circ. Res.* **122**, 1661–1674 (2018).
11. A. Zernecke *et al.*, *Circ. Res.* **127**, 402–426 (2020).
12. S. McArdle *et al.*, *Circ. Res.* **125**, 1038–1051 (2019).
13. R. C. Araneda, Z. Peterlin, X. Zhang, A. Chesler, S. Firestein, *J. Physiol.* **555**, 743–756 (2004).
14. Y. Li *et al.*, *ACS Chem. Biol.* **9**, 2563–2571 (2014).
15. R. C. Araneda, A. D. Kini, S. Firestein, *Nat. Neurosci.* **3**, 1248–1255 (2000).
16. W. B. Rizzo, *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **1841**, 377–389 (2014).
17. S. Feil *et al.*, *Circ. Res.* **115**, 662–667 (2014).
18. M. R. Bennett, S. Sinha, G. K. Owens, *Circ. Res.* **118**, 692–702 (2016).
19. T. Bozza, P. Feinstein, C. Zheng, P. Mombaerts, *J. Neurosci.* **22**, 3033–3043 (2002).
20. L. Folkersen *et al.*, *Mol. Med.* **18**, 669–675 (2012).
21. Y. I. Miller, *Future Cardiol.* **1**, 785–792 (2005).
22. V. Camarena *et al.*, *eLife* **6**, e29750 (2017).
23. J. R. Yaron *et al.*, *Cell Death Dis.* **6**, e1954 (2015).
24. K. V. Swanson, M. Deng, J. P. Ting, *Nat. Rev. Immunol.* **19**, 477–489 (2019).
25. K. Tsuchiya *et al.*, *Cell Rep.* **34**, 108887 (2021).
26. R. C. Coll *et al.*, *Nat. Med.* **21**, 248–255 (2015).
27. B. A. McKenzie *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6065–E6074 (2018).
28. J. J. Hu *et al.*, *Nat. Immunol.* **21**, 736–745 (2020).
29. H. Esterbauer, G. Wäg, H. Puhl, *Br. Med. Bull.* **49**, 566–576 (1993).
30. M. T. R. Morales, J. J. Rios, R. Aparicio, *J. Agric. Food Chem.* **45**, 2666–2673 (1997).
31. L. Xu, X. Yu, M. Li, J. Chen, X. Wang, *Int. J. Food Prop.* **20** (suppl. 3), S2926–S2938 (2017).
32. P. Mukherjee *et al.*, *J. Lipid Res.* **59**, 1818–1840 (2018).
33. J. J. Piotrowski, G. C. Hunter, C. D. Eskelson, M. A. Dubick, V. M. Bernhard, *Life Sci.* **46**, 715–721 (1990).
34. A. Daugherty *et al.*, *Arterioscler. Thromb. Vasc. Biol.* **37**, e131–e157 (2017).
35. R. Sharma *et al.*, *eLife* **6**, e21895 (2017).

CELL BIOLOGY

Bacterial gasdermins reveal an ancient mechanism of cell death

Alex G. Johnson[1,2]†, Tanita Wein[3]†, Megan L. Mayer[4], Brianna Duncan-Lowey[1,2], Erez Yirmiya[3], Yaara Oppenheimer-Shaanan[3], Gil Amitai[3], Rotem Sorek[3]*, Philip J. Kranzusch[1,2,5]*

Gasdermin proteins form large membrane pores in human cells that release immune cytokines and induce lytic cell death. Gasdermin pore formation is triggered by caspase-mediated cleavage during inflammasome signaling and is critical for defense against pathogens and cancer. We discovered gasdermin homologs encoded in bacteria that defended against phages and executed cell death. Structures of bacterial gasdermins revealed a conserved pore-forming domain that was stabilized in the inactive state with a buried lipid modification. Bacterial gasdermins were activated by dedicated caspase-like proteases that catalyzed site-specific cleavage and the removal of an inhibitory C-terminal peptide. Release of autoinhibition induced the assembly of large and heterogeneous pores that disrupted membrane integrity. Thus, pyroptosis is an ancient form of regulated cell death shared between bacteria and animals.

I n mammals, gasdermin proteins execute pyroptotic cell death by oligomerizing into membrane pores that release inflammatory cytokines and induce cell lysis. The human genome encodes six gasdermin proteins (GSDMA to GSDME and pejvakin), including the prototypical member GSDMD (*1–3*). Gasdermin activation requires caspase- or granzyme-mediated cleavage of an interdomain linker that liberates a lipophilic N-terminal domain (NTD) from a large inhibitory C-terminal domain (CTD) (*4–6*). Proteolysis enables gasdermin NTD oligomerization and the formation of membrane pores important for innate immunity in mammals and primitive eukaryotes (*7–9*). Recent structural analyses have explained a mechanism of gasdermin pore formation (*5, 6, 10, 11*), but the evolutionary origin and biological roles of diverse gasdermin proteins remain unknown (*12*).

While analyzing bacterial antiphage defense islands, we identified uncharacterized genes with predicted homology to mammalian gasdermins (table S1). Sequence analysis revealed 50 bacterial gasdermins (bGSDMs) that form a clade distinct from that of eukaryotic homologs (Fig. 1A and fig. S1) (*7, 9, 13*). We determined crystal structures of bGSDMs from *Bradyrhizobium tropiciagri* and *Vitiosangium sp.*, which revealed that bGSDMs each adopt a shared overall architecture that exhibits notable homology to the mammalian gasdermin NTD (fig. S2B and table S2), including the conservation of a twisted central antiparallel β sheet and the shared placement of connecting helices and strands throughout the periphery (Fig. 1, B and C).

The structures revealed complete absence of the large α-helical CTD required to maintain mammalian gasdermins in an autoinhibited state (Fig. 1). Although lacking the CTD, the bGSDM structures adopted the same conformation as that of the inactive mammalian gasdermin complex (Fig. 1, B and C). In the inactive structure of mammalian GSDMA3, the NTD forms two interfaces with the CTD that mediate autoinhibition, with the primary interface at the α1 helix and the β1-β2 hairpin (Fig. 1C) (*5, 11*). Cleavage of GSDMA3 results in NTD activation through the lengthening of strands β3, β5, β7, and β9 and oligomerization of ~27 protomers into a membrane-spanning pore (*2, 5, 6*). Both the *Bradyrhizobium* and

Vitiosangium bGSDM structures contained strands equivalent to GSDMA3 β1 to β2 and β6 to β9, but in bGSDMs, a short C-terminal peptide wrapped around the twisted β sheet core and terminated across strand β2 to stabilize the inactivated state (Fig. 1, B and C).

While building the bGSDM atomic models, we observed a snakelike density protruding from the *Bradyrhizobium* cysteine C3 side-chain. The density occupies a hydrophobic tunnel across the protein that is capped by F25 from the C-terminal peptide. In the 1.5-Å *Bradyrhizobium* bGSDM electron density map, the density could be assigned as a 16-carbon palmitoyl thioester (Fig. 1D and fig. S2C). We confirmed bGSDM palmitoylation with mass spectrometry and found that a cysteine at this position is conserved in gasdermins across most bacteria and some fungi (Fig. 1A and fig. S3, A and B). The presence of the palmitoyl in a hydrophobic cavity suggests that bGSDM palmitoylation occurs through autocatalysis (*14*). Palmitoylation contributes to stability of the inactive state protein (Fig. 1E), and modeling suggests substantial reorganization of residues along the hydrophobic tunnel during bGSDM activation (Fig. 1D and fig. S2, C and D) (*6*).

The majority of bGSDMs (43 of 50) are genomically encoded next to one or more genes with a predicted protease domain (Fig. 2A; fig. S5, A to C; and table S1). In most cases, the associated proteases are caspase-like peptidases, including peptidase C14 (Pfam database ID PF00656) and CHAT (Pfam ID PF12770) proteases (Fig. 2B and fig. S5A). Fungal gasdermins are also commonly encoded next to protease domain–containing genes (40 of 52) (table S3 and fig. S5B) and are activated through proteolysis (*13*). bGSDM-protease systems are found in diverse bacteria and archaea, as well as in metagenomic samples of prokaryotic origin (fig. S5D and table S4). Analysis of the bGSDM-associated proteases revealed that they are fused to divergent repeat or NACHT domains frequently involved in pathogen recognition and inflammasome function in human innate immunity (Fig. 2B and fig. S5C) (*15*). bGSDM genes are occasionally encoded near known immune defense systems (Fig. 2A, fig. S7A, and tables S1 and S4), so we tested bGSDM systems for a role in antiphage defense. bGSDM systems evolutionarily distant from the model organism *Escherichia coli* exhibited no discernible phage restriction (fig. S6). However, a

[1]Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA. [2]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02115, USA. [3]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. [4]Harvard Center for Cryo-Electron Microscopy, Harvard Medical School, Boston, MA 02115, USA. [5]Parker Institute for Cancer Immunotherapy, Dana-Farber Cancer Institute, Boston, MA 02115, USA.
*Corresponding author. Email: philip_kranzusch@dfci.harvard.edu (P.J.K.); rotem.sorek@weizmann.ac.il (R.S.)
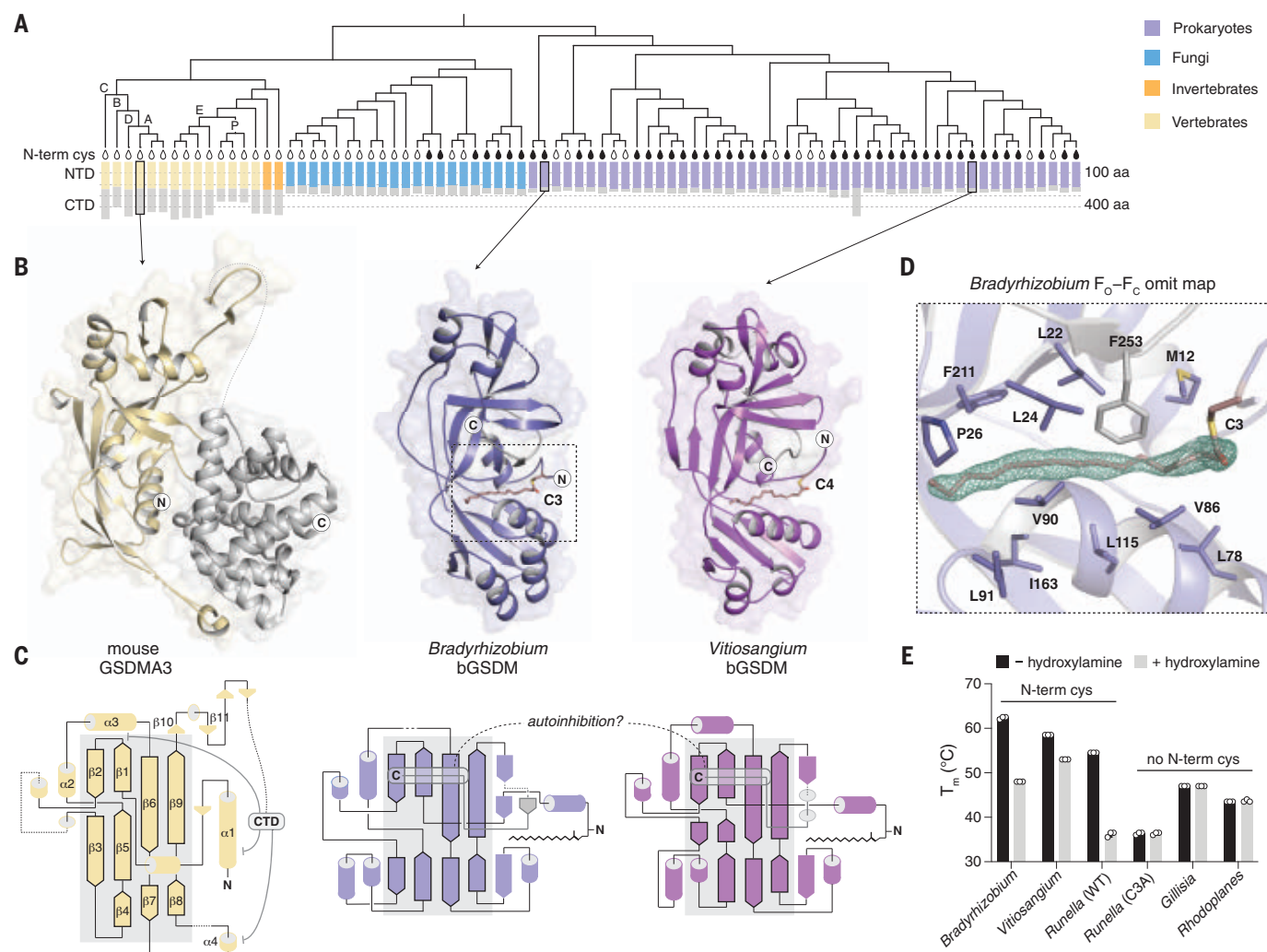†These authors contributed equally to this work.

Fig. 1. Structures of bGSDMs reveal homology with mammalian cell death effectors. (**A**) Gasdermin phylogenetic tree. The sizes of the gasdermin NTDs and CTDs are depicted. Vertebrate gasdermins are labeled with single letters ("A" to "E") indicating human gasdermins GSDMA to GSDME, and "P" depicts pejvakin. The black teardrop indicates a conserved N-terminal cysteine (N-term cys). A representative set of 20 fungal gasdermins are included in the tree. aa, amino acid. (**B**) Crystal structures of bGSDMs from species of the genera *Bradyrhizobium* and *Vitiosangium*. bGSDM structures reveal homology to the NTD of mammalian gasdermins in an inactive conformation, including mouse GSDMA3 (Protein Data Bank ID 5B5R). (**C**) Gasdermin topology diagrams indicate a conserved central core of the bacterial and mammalian NTD. bGSDMs notably lack the CTD required

for autoinhibition of mammalian gasdermins and instead encode a short C-terminal peptide. (**D**) Simulated annealing F_O–F_C omit map (contoured at 3.0 σ) from the *Bradyrhizobium* bGSDM fit with a palmitoyl modification at C3. Omit map is shown as green mesh, and select residues forming a hydrophobic pocket around the palmitoyl group are indicated. (**E**) Melting temperatures (T_m) of bGSDMs with and without N-terminal cysteines, as determined with thermofluor assays. Data are the means and standard deviations of three technical replicates and are representative of three independent experiments. WT, wild-type. 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.

four-gene operon from *Lysobacter enzymogenes* exhibited robust defense against coliphages T4, T5, and T6 (Fig. 2, C and D, and fig. S6, B and C). Deletion of the bGSDM gene from the *Lysobacter* operon abolished protection (Fig. 2, C and D, and fig. S6C). Thus, the bGSDM is essential for defense.

Expression of some of the bGSDM-protease systems in *E. coli* induced potent cellular toxicity in the absence of phage infection (Fig. 2,

E and F; fig. S7, B and C; and table S5). Particularly strong toxicity was observed for a *Runella* system, which required bGSDM palmitoylation (Fig. 2E and fig. S7, B and C). Time-lapse microscopy in the presence of propidium iodide (PI) showed that cells expressing the *Runella* system ceased dividing and lost membrane integrity, which suggests that bGSDM activation induces membrane disruption (Fig. 2, E and F; fig. S7D; and movies S1 and S2). Mutation of

the predicted *Runella* caspase-like protease active site residues H796 and C804 ablated all cellular toxicity (Fig. 3A). The *Runella* bGSDM and protease only induced cellular toxicity when expressed together, which suggests that the protease targets bGSDM during system activation (Fig. 3A and fig. S8A). In fact, a mutation that disrupted the active site of the second trypsin-like protease in the *Lysobacter* bGSDM system abolished antiphage defense (fig. S8B).

Fig. 2. bGSDMs are associated with proteases, defend from phages, and execute cell death.

(**A**) Representative instances of bGSDMs and associated proteases in their genomic neighborhoods. Genes known to be involved in antiphage defense are shown in yellow. TA, toxin-antitoxin; Abi, abortive infection; ATPase, adenosine triphosphatase. (**B**) Types of proteases found adjacent to bGSDMs (n = 59). Some bGSDMs appear with more than one adjacent protease. Caspase-like proteases include peptidase C14 (n = 15) and CHAT (n = 23). Cases in which the protease gene also encodes an additional domain are indicated. TPR, tetratricopeptide repeat; LRR, leucine-rich repeat. (**C**) A bGSDM-containing operon protects against phages. The efficiency of plating of phages on *E. coli* MG1655 cells expressing the *Lysobacter* bGSDM WT or mutated operon is shown. Data represent plaque-forming units (PFU) per milliliter and are the averages of three independent replicates, with individual data points overlaid. GFP represents a control strain. Statistical significance was determined by a one-way analysis of variance (ANOVA) and Tukey multiple comparison test. Not significant (n.s.) ≥ 0.05; **P = 0.001 to 0.01. (**D**) Growth of liquid cultures of *E. coli* expressing the WT and mutated *Lysobacter* bGSDM operons. Cells were infected with phage T6. For each experiment, data represent one out of three biological replicates (replicates are shown in fig. S6). OD$_{600}$, optical density at 600 nm. (**E**) The *Runella* bGSDM operon causes cell death. *E. coli* DH5α cells expressing the *Runella* protease and WT or C3A-mutated bGSDM were examined by time-lapse microscopy. Overlay images from PI (red) and phase contrast of cells captured at the start of the experiment and after 120 min of incubation are shown. Scale bar, 2 μm. (**F**) bGSDM operons are toxic. Cells encoding protease and WT or mutated bGDSM were plated in 10-fold serial dilution on LB-agar in conditions that repress operon expression (1% glucose) or induce expression (0.2% arabinose).
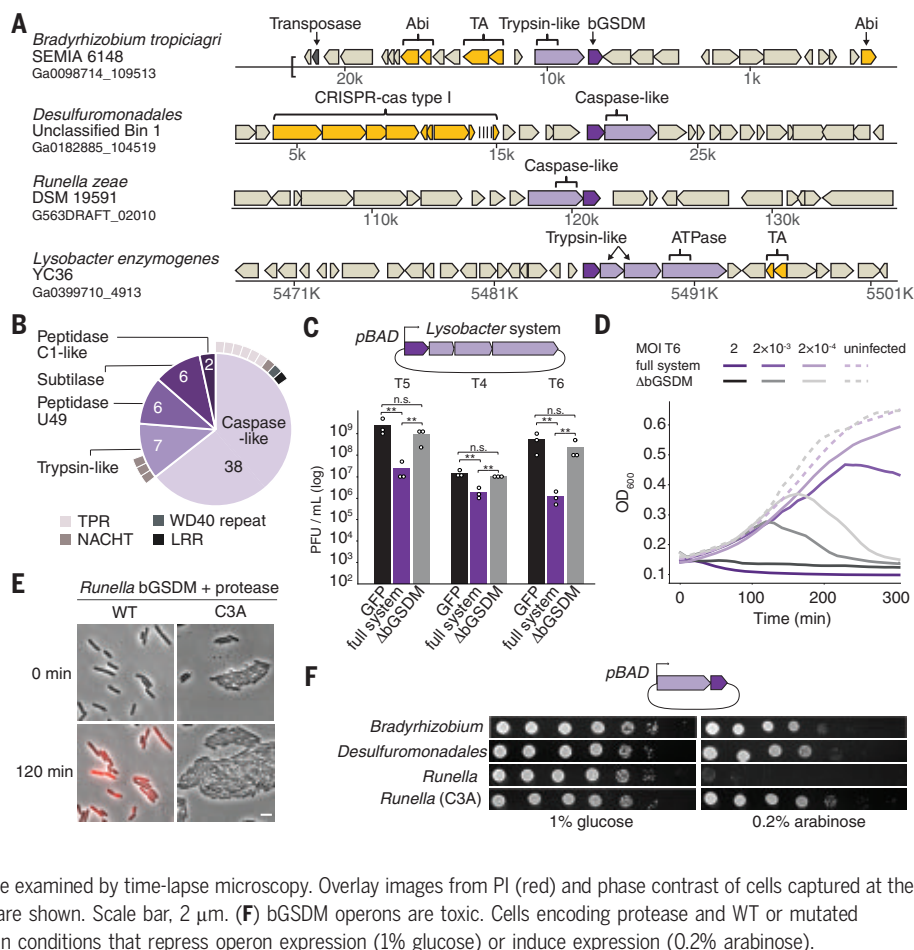
Fig. 3. bGSDMs are activated by proteolytic cleavage.

(**A**) Toxicity of *Runella* bGSDM in vivo requires the associated protease. Bacteria expressing WT and mutated versions of the *Runella* bGSDM–protease operon were grown on LB-agar in conditions that repress or induce expression. Data represent colony-forming units (CFU) per milliliter, and bar graphs represent an average of three independent replicates, with individual data points overlaid. Asterisks indicate statistically significant differences compared with the respective noninduced control using two-sided t test. n.s. ≥ 0.05; ***P = 0.0001 to 0.001; ****P < 0.0001. (**B**) *Runella* bGSDM cleavage by its associated protease is dependent on catalytic histidine and cysteine residues in vitro. Fifteen percent SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gels were run after cleavage at room temperature for 18 hours and visualized by Coomassie staining. (**C**) The *Runella* bGSDM crystal structure and protease cleavage site. The *Runella* bGSDM structure is shown in lavender with the last 21 amino acids highlighted as gray spheres. (**D**) Close-up view of the *Runella* bGSDM cleavage site wherein cleavage occurs after the P1 L247 residue. (**E**) Structural overview of the *Bradyrhizobium* CTD and autoinhibitory interactions. The bGSDM is colored purple except for its last 16 residues, which are colored yellow. Insets show interactions of F245 and F247 adjacent to D21 of the N-terminal β sheet (top) and F253 and the palmitoyl modification at C3 (bottom). The 2F$_O$–F$_C$ (contoured at 1.5 σ) map is shown as gray mesh fit to the last 16 residues.
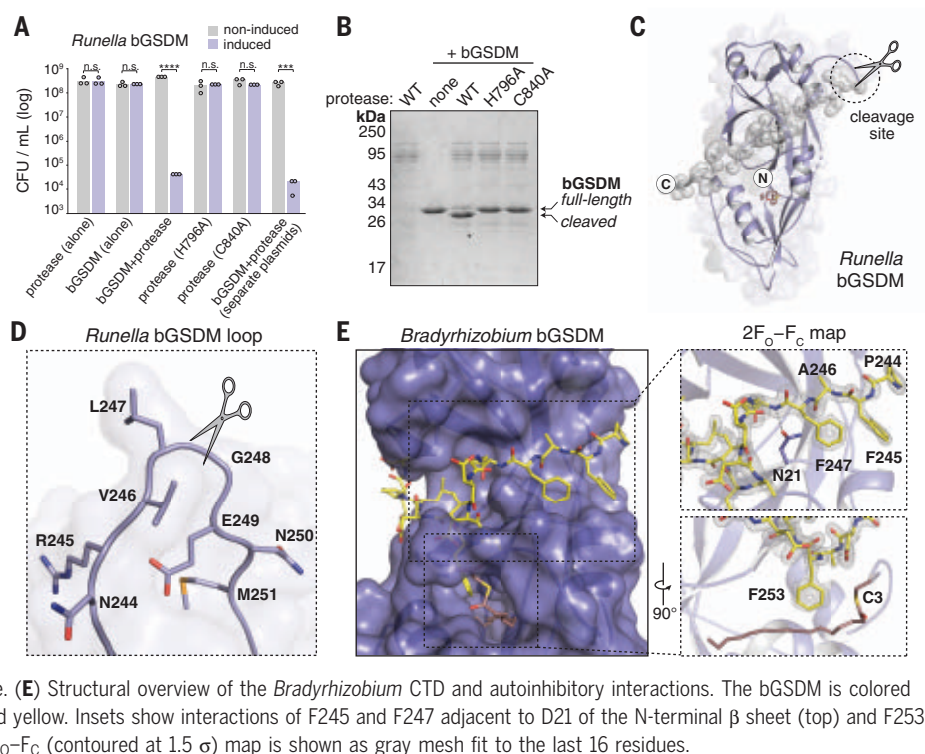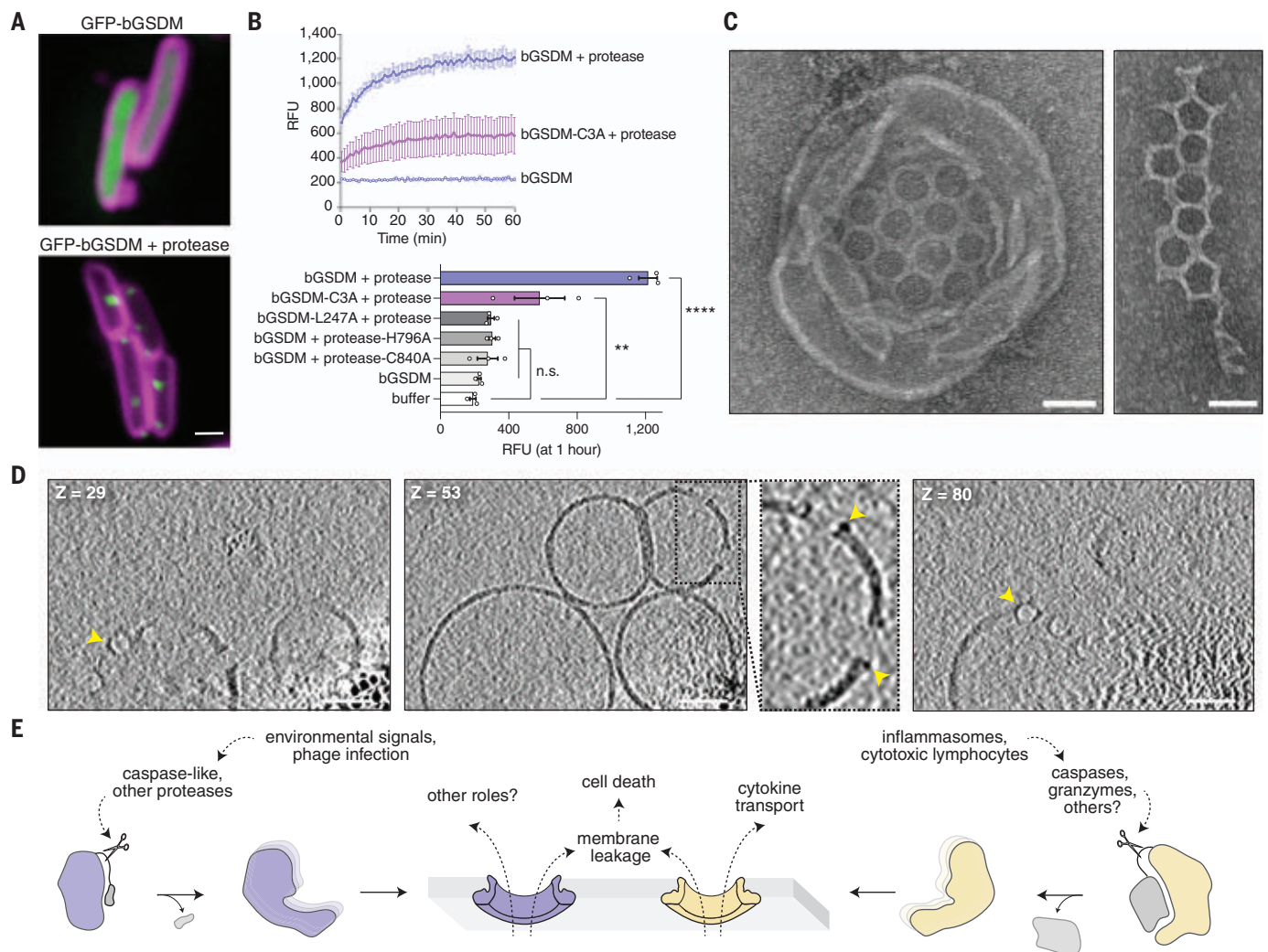
Fig. 4. Cleaved bGSDMs form membrane pores to elicit cell death.
(**A**) GFP was fused to the N terminus of the *Runella* bGSDM. Cells expressing GFP-bGSDM alone (top) or with the caspase-like protease (bottom) are shown. GFP is colored green. Membrane dye (FM4-64) is in magenta. Scale bar, 1 μm. (**B**) Cleaved *Runella* gasdermin permeabilizes liposome membranes. Relative fluorescence units (RFU) were measured continuously from cleavage reactions of dioleoylphosphatidylcholine (DOPC) liposomes loaded with TbCl3 with an external solution containing 20 μM dipicolonic acid (DPA). The top plot represents an example of time-course liposome leakage, whereas the bottom bar chart shows values for each condition at 60 min. Error bars represent the SEM of three technical replicates, and statistical significance was determined by one-way ANOVA and Tukey multiple comparison test. n.s. ≥ 0.05; **P = 0.001 to 0.01; ****P < 0.0001. (**C**) Negative stain electron microscopy of *Runella* gasdermin pores in DOPC liposomes (left) and in mesh-like arrays (right). Scale bars, 50 nm. (**D**) Slices from representative tomogram (1 of 10) of *Runella* gasdermin pores in DOPC liposomes, at three different depths (Z). Yellow arrowheads indicate pores inserted within the liposome membrane. Scale bars, 50 nm. (**E**) Model of pyroptosis for bGSDMs and mammalian gasdermins.

We focused on the *Runella* system and reconstituted cleavage with purified components (Fig. 3B and fig. S8, A to F). Coincubation with the protease resulted in specific bGSDM cleavage and formation of a lower–molecular weight *Runella* bGSDM species (fig. S8A). Cleavage requires the protease catalytic residues but not bGSDM palmitoylation (Fig. 3B and fig. S8, D and E). Using mass spectrometry, we determined that the *Runella* bGSDM cleavage site occurs after the P1 residue L247 (fig. S9, A and B). A 2.9-Å structure of the *Runella* bGSDM (table S2) revealed

that cleavage occurs in a loop that immediately precedes the C-terminal peptide (Fig. 3, C and D). Packing in the *Runella* bGSDM crystal lattice additionally indicates an ability of the peptide to dissociate from the bGSDM face, which supports release after cleavage (fig. S10A).

Analysis of the high-resolution *Bradyrhizobium* bGSDM structure explains how the C-terminal peptide restrains the bGSDM core (Fig. 3E). *Bradyrhizobium* bGSDM F245 and F247 lay along the surface formed between the mammalian β9 strand and the α1 helix equivalent

positions and are further supported with contacts between N21 and the peptide backbone. A *Bradyrhizobium*-specific β strand from N21 to L24 extends off the β9 strand equivalent and is supported by a short parallel β strand from F253 to D255. *Bradyrhizobium* bGSDM F253 latches over the palmitoyl modification, with similar hydrophobic contacts also observed in the *Runella* and *Vitiosangium* structures (fig. S10, A and B). The *Bradyrhizobium* bGSDM C-terminal peptide terminates below the strand equivalent to β2 and is supported by hydrogen bonds from R27 to the

L256 backbone and N29 to E258. Truncation of the C-terminal peptide in the *Runella*, *Bradyrhizobium*, or *Vitiosangium* bGSDM constructs led to arrested cell growth, which confirms that the C-terminal peptide is required to maintain the bGSDM autoinhibition (fig. S11, A and B).

We next used mutagenesis of the *Runella* bGSDM system to define the specificity of proteolytic cleavage and bGSDM activation. In vitro, the L247 P1 position was essential for cleavage, and proteolysis was inhibited by mutations that disrupt the P1′ glycine and the P4, P3, P2, and P3′ residues (fig. S11, C and D). Likewise, mutations that disrupt the P1 and P1′ positions eliminated toxicity in vivo (fig. S11, E and F). The *Runella* protease was not capable of activating divergent bGSDMs engineered to contain the *Runella* cleavage loop, which suggests that additional contacts specify bGSDM recognition (fig. S11, G and H). Thus, like their mammalian homologs, bGSDMs are cell death effectors activated by proteolytic cleavage.

To determine whether activated bGSDMs associate with bacterial membranes, we fused green fluorescent protein (GFP) to the N terminus of the *Runella* bGSDM and visualized expression in *E. coli*. Upon coexpression with the *Runella* protease, GFP-bGSDM coalesced into membrane-associated puncta and induced cellular toxicity (Fig. 4A and fig. S12, A to C). Transmission electron microscopy analysis of *E. coli* expressing the active *Runella* bGSDM system revealed clear disruption of membrane integrity (fig. S13, A to C). In vitro reconstituted *Runella* bGSDM activity demonstrated that cleaved *Runella* bGSDMs permeabilized liposomes and caused rapid release of the internal contents (Fig. 4B and fig. S14, A and B). Protease active-site or bGSDM cleavage-site mutations disrupted all liposome permeabilization, which confirms that proteolysis is essential for bGSDM activation (Fig. 4B and fig. S14B). Blocking bGSDM palmitoylation with mutation of residue C3 reduced but did not abolish liposome leakage or membrane-associated puncta formation in cells (Fig. 4B and fig. S12A). Likewise, a C7A mutation to the putative *Lysobacter* bGSDM palmitoylation site was not sufficient to abolish antiphage defense (fig. S8B), which suggests that lipid-modification supports but is not required for membrane permeabilization.

To compare the bGSDM pore with its mammalian counterparts, we used electron microscopy to visualize *Runella* bGSDM cleavage reactions and liposomes (Fig. 4C and fig. S15, A to C). bGSDM pores were observed within liposomes and as fragmented mesh-like arrays. Cryo–electron microscopy (cryo-EM) and two-dimensional (2D) classification analysis of detergent-solubilized complexes revealed that *Runella* bGSDM forms a ringlike pore that exhibits a width of ~50 Å and an inner diameter ranging from 200 to 300 Å (fig. S17, A to D). *Runella* bGSDM pores within liposomes measured ~240 to 330 Å—larger than the 135- to 215-Å mammalian gasdermin pores (fig. S17, A to D) (*6, 10, 16*). We also reconstituted cleavage of a bGSDM from a metagenomic *Bacteroidetes* scaffold and observed smaller 130- to 190-Å pores within liposomes, which suggests heterogeneity in the architecture of diverse bGSDM pores (fig. S18, A to D, and fig. S19, A to C). Cryo–electron tomography (cryo-ET) tilt series reconstructions of the pore-liposome assemblies confirmed that bGSDM pores span the liposomal surface to disrupt membrane integrity (Fig. 4D, fig. S20, and movies S3 and S4).

Our results support a model for gasdermin pore formation and effector function that has notable parallels between bacteria and mammals (Fig. 4E). bGSDM systems can exert antiphage defense, and the fusion of bGSDM-associated proteases with NACHT and repeat domains suggests that, similar to inflammasome sensors in mammals, foreign pathogen recognition may control the initiation of gasdermin cleavage (Fig. 2, B to D) (*15, 17*). In both mammalian gasdermin and bGSDM systems, proteolytic cleavage after the lipophilic NTD releases gasdermin inhibition. The notably short C-terminal peptide responsible for bGSDM inhibition suggests the possibility that short-form eukaryotic gasdermins, including pejvakin, may undergo activation through a similar mechanism. Furthermore, widespread palmitoylation of bGSDMs indicates that cysteine modifications are a conserved mechanism for regulating gasdermin pores (*18*). The size distribution of pores from *Runella* and *Bacteroidetes* species might suggest that bGSDM pores, like those in mammals, could be customized for the secretion of certain molecules (*10*). Defining the cues that activate bGSDM systems will provide insight into their roles in prokaryotic biology and the origins of pyroptotic cell death.

REFERENCES AND NOTES

1. N. Kayagaki et al., *Nature* **526**, 666–671 (2015).
2. J. Shi et al., *Nature* **526**, 660–665 (2015).
3. W. T. He et al., *Cell Res.* **25**, 1285–1298 (2015).
4. X. Liu et al., *Nature* **535**, 153–158 (2016).
5. J. Ding et al., *Nature* **535**, 111–116 (2016).
6. J. Ruan, S. Xia, X. Liu, J. Lieberman, H. Wu, *Nature* **557**, 62–67 (2018).
7. S. Jiang, Z. Zhou, Y. Sun, T. Zhang, L. Sun, *Sci. Immunol.* **5**, eabd2591 (2020).
8. J. Lieberman, H. Wu, J. C. Kagan, *Sci. Immunol.* **4**, eaav1447 (2019).
9. A. Daskalov, P. S. Mitchell, A. Sandstrom, R. E. Vance, N. L. Glass, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18600–18607 (2020).
10. S. Xia et al., *Nature* **593**, 607–611 (2021).
11. Z. Liu et al., *Immunity* **51**, 43–49.e4 (2019).
12. P. Broz, P. Pelegrín, F. Shao, *Nat. Rev. Immunol.* **20**, 143–157 (2020).
13. C. Clavé et al., *bioRxiv* 2021.06.03.446900 [Preprint] (2021).
14. D. Kümmel, U. Heinemann, M. Veit, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12701–12706 (2006).
15. M. Shi, P. Zhang, S. M. Vora, H. Wu, *Curr. Opin. Cell Biol.* **63**, 194–203 (2020).
16. J. M. Hansen et al., *Cell* **184**, 3178–3191.e18 (2021).
17. G. Kaur, A. M. Burroughs, L. M. Iyer, L. Aravind, *eLife* **9**, e52696 (2020).
18. F. Humphries et al., *Science* **369**, 1633–1637 (2020).

PHYSICS

Observation of a gravitational Aharonov-Bohm effect

Chris Overstreet[1]†, Peter Asenbaum[1,2]†, Joseph Curti[1], Minjeong Kim[1], Mark A. Kasevich[1]*

Gravity curves space and time. This can lead to proper time differences between freely falling, nonlocal trajectories. A spatial superposition of a massive particle is predicted to be sensitive to this effect. We measure the gravitational phase shift induced in a matter-wave interferometer by a kilogram-scale source mass close to one of the wave packets. Deflections of each interferometer arm due to the source mass are independently measured. The phase shift deviates from the deflection-induced phase contribution, as predicted by quantum mechanics. In addition, the observed scaling of the phase shift is consistent with Heisenberg's error-disturbance relation. These results show that gravity creates Aharonov-Bohm phase shifts analogous to those produced by electromagnetic interactions.

I n classical physics, the state of a particle is given by its position and momentum. Because the trajectory of a classical particle is determined by its interactions with local fields, the deflection of a particle can be used to observe a field. However, a classical particle cannot measure the action along its trajectory.

The situation is different in quantum mechanics. As Aharonov and Bohm argued in 1959, a particle in a spatial superposition is sensitive to the potential energy difference between its wave packets even if the field vanishes along their trajectories (1). A matter-wave interferometer can therefore measure a phase shift due to the potential even if the interferometer arms are not deflected. This phase shift ϕ_{AB} is given by the action difference ΔS between arms according to the expression $\phi_{AB} = \Delta S/\hbar$ (1). The Aharonov-Bohm effect can be described in terms of a quantum particle interacting with a classical electromagnetic potential (1) or in terms of a quantum particle interacting locally with a quantized electromagnetic field and source (2).

The Aharonov-Bohm effect induced by a magnetic field was first observed in 1960 (3). Since then, experiments have identified related effects in a variety of systems (4, 5). The successful observation of Aharonov-Bohm phase shifts in the electromagnetic domain raises a question: Can analogous phase shifts be caused by gravity as well? Quantum mechanics predicts that gravity can create an action difference between interferometer arms, giving rise to a "gravitational Aharonov-Bohm effect" (6). In general relativity, this phenomenon is described by the gravitationally induced proper time difference between the geodesics corresponding to the interferometer arm trajectories. This effect has not previously been observed. Its experimental detection in an atom interferometer was proposed in (7).

Prior experiments (8) were not sensitive to the gravitational Aharonov-Bohm effect because $\Delta S \approx 0$ when the wave packet separation is small compared to the length scale of the gravitational potential (9, 10). The interferometer phase in this regime is proportional to the deflection of the atomic wave packet with respect to its beam splitters (11, 12) and is independent of the particle mass m. However, when the wave packet separation is large, ΔS becomes nonzero. Qualitatively, an interferometer enters this nonlocal regime when the wave packet separation becomes larger than the distance between the source mass and an interferometer arm.

We use a light-pulse ^{87}Rb atom interferometer (12) with large–momentum-transfer beam splitters ($52\hbar k$, where k is the laser wave number) and large wave packet separation (25 cm) to measure the phase shift induced by a tungsten source mass. At its closest approach, one interferometer arm passes within 7.5 cm of the source mass, which alters its proper time (Fig. 1A). The source mass also deflects the interferometer arms. To quantify the influence of deflections on the phase shift, we measure the deflections with a pair of $4\hbar k$ interferometers (2-cm wave packet separation). The phase shift of the $52\hbar k$ interferometer deviates strongly from the deflection-induced phase contribution. We show that $\phi_{AB} \neq 0$, demonstrating the gravitational Aharonov-Bohm effect in this system.

In the experiment (13, 14), a cloud of ^{87}Rb is evaporatively cooled to ~1 μK in a magnetic trap, magnetically lensed to a velocity width of 2 mm/s, and launched into a 10-m vacuum chamber at 13 m/s by an optical lattice. The lattice depth is decreased for a short interval during the launch to release half of the atoms at a lower velocity [see materials and methods for details (15)]. After the launch, the two clouds are decelerated to a relative momentum of $2\hbar k$ by sequential Bragg transitions and are used as the inputs of a single-source gradiometer (16) with baseline 24 cm (Fig. 1B). The matter-wave beam splitters and mirrors consist of laser pulses that transfer momentum to the atoms via Bragg transitions. The midpoint trajectory of each $4\hbar k$ interferometer is matched to the trajectory of one arm of the $52\hbar k$ interferometer. The $52\hbar k$, upper $4\hbar k$, and lower $4\hbar k$ gradiometers are implemented in separate shots. The upper interferometer in each gradiometer is sensitive to the source mass, whereas the lower interferometer mainly acts as a phase reference. This reference is necessary to remove contributions to the phase shift arising from fluctuations in the phase of the optical field. The time between the initial beam splitter pulse and the mirror pulse (interferometer time T) is
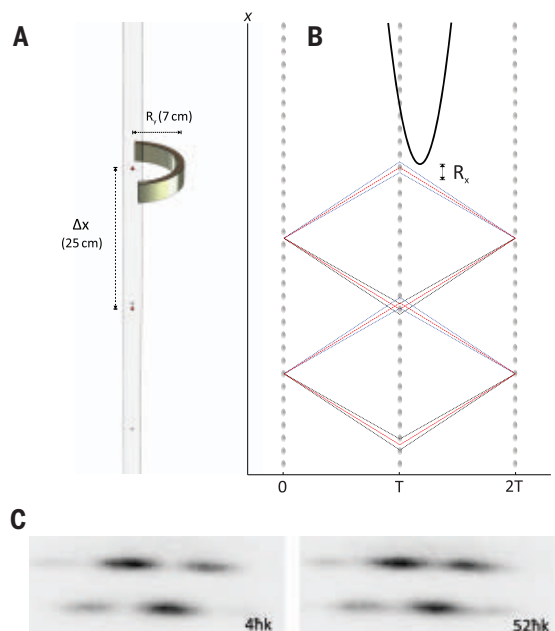
[1]Department of Physics, Stanford University, Stanford, CA 94305, USA. [2]Institute for Quantum Optics and Quantum Information (IQOQI) Vienna, Austrian Academy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria.
*Corresponding author. Email: kasevich@stanford.edu
†These authors contributed equally to this work.

Fig. 1. Experimental setup.
(**A**) Interferometer arms, tungsten source mass, and laser beam splitter. One arm of a light-pulse atom interferometer approaches the source mass, while the other arm remains far away. (**B**) Space-time diagram of gradiometer geometries in a freely falling reference frame. The red, blue, and black dotted lines represent the trajectories of the $52\hbar k$, upper $4\hbar k$, and lower $4\hbar k$ gradiometers, respectively, while the solid black line represents the trajectory of the source mass. Interferometer pulses (gray dashed lines) occur at times $t = 0$, $t = T$, and $t = 2T$. (**C**) Fluorescence images of interferometer output ports, $4\hbar k$ (left) and $52\hbar k$ (right).

0.82 s. After a gradiometer sequence is complete, the atoms are imaged by resonant scattering (Fig. 1C). The output ports of the two interferometers are imaged simultaneously on separate, vertically displaced cameras. A horizontal detection fringe (13) is applied to improve phase readout. As in (13), the direction of the detection fringe is reversed to suppress imaging-related systematic effects.

The source mass is a 170° semicircular ring with inner radius $R_y = 6.8$ cm, outer radius 7.8 cm, and height 3 cm, chosen to be consistent with apparatus geometric constraints. The ring has a mass of 1.25 kg and is 99.95% pure tungsten. We verified that the source mass is nonmagnetic at the level required for this work [see materials and methods for details (15)]. The mass is placed within the magnetic shield that surrounds the interferometry region at a height of 27 cm below the end cap. This position corresponds to $R_x = 0$, where R_x is the vertical displacement between the source mass and the apex of the upper arm trajectory.

In the phase shift plots of the $52\hbar k$ and $4\hbar k$ gradiometers as a function of R_x (Fig. 2), each data point represents the difference in the gradiometer phase with and without the source mass installed. This differential measurement technique suppresses the phase contribution from Earth's gravity gradient along with other systematic effects that are common to the two configurations.

The $4\hbar k$ interferometers have a small wave packet separation and can be understood as deflection measurements. This property enables a simple explanation of the shapes of the curves in Fig. 2B. At large approach distances ($R_x < 0$, $|R_x| \gg R_y$), all interferometers are far below the source mass, and the phase shifts are small. As the launch height is increased, the upper $4\hbar k$ interferometer approaches the source mass. The atoms are deflected upward toward the source mass, making the phase shift more negative. When $R_x > 0$, the atoms of the upper $4\hbar k$ interferometer spend time above the source mass and begin to be deflected downward by it; the phase shift thus passes through zero near $R_x = 4$ cm and becomes positive. At these approach distances, the atoms of the lower $4\hbar k$ interferometer begin to be deflected upward by the source mass, and the phase shift becomes negative. Unlike the upper $4\hbar k$ interferometer, the phase shift of the $52\hbar k$ interferometer remains negative for all values of R_x, indicating that it cannot be explained solely by deflections. The $52\hbar k$ gradiometer phase uncertainty in a single shot is typically about 30 mrad, inferred from the observed standard deviation of a sequence of shots.

To interpret our measurement as an Aharonov-Bohm experiment, we characterize the relationship between deflections, action differences,

and the interferometer phase. The phase ϕ of a light-pulse matter-wave interferometer in a gravitational potential can be written as the sum of two terms: $\phi = \phi_{MP} + \phi_{\Delta S}$ (12). These terms have distinct physical interpretations. The "midpoint phase," ϕ_{MP}, arises from local atom-light interactions during beam splitter and mirror pulses (11, 12). For a Bragg interferometer, the midpoint phase is expressed as $\phi_{MP} = -\sum_i k_i \cdot \bar{x}_i$, where k_i is the wave number difference applied to the two arms by the ith light pulse and $\bar{x}_i$ is the midpoint displacement of the interferometer arms at the ith light pulse with respect to the optical phase reference. Classically, the midpoint phase could be measured by observing the positions of particles that travel along the interferometer arm trajectories. By contrast, the beyond-midpoint phase is given by

$$\phi_{\Delta S} = \frac{\Delta S}{\hbar} = \frac{m}{\hbar} \int \left(\left[V(x_1, t) - V(x_2, t) \right] - \frac{\Delta x}{2} \left[\frac{\partial V(x_1, t)}{\partial x} + \frac{\partial V(x_2, t)}{\partial x} \right] \right) dt \quad (1)$$

for gravitational potential V, wave packet separation Δx, and arm trajectories $x_1(t)$, $x_2(t)$. The first term in the integrand depends on the potential energy difference between arms, whereas the second term depends on the kinetic energy difference. This phase is pro-

portional to the proper time evolved around a closed interferometer loop. In principle, $\phi_{\Delta S}$ could be measured by observing the phase difference of two clocks with frequency mc^2/h that travel along the interferometer arms (17), but $\phi_{\Delta S}$ cannot be inferred by observing the interferometer arm trajectories. In previous gravitational measurements (16, 18) and in our $4\hbar k$ gradiometers, $\phi_{\Delta S}$ is smaller than the measurement resolution.

In an ideal Aharonov-Bohm measurement, the interferometer arm trajectories would be completely unaltered by the potential. In that case, we would have $\frac{\partial V}{\partial x} = 0$ along both trajectories, $\phi_{MP} = 0$, and $\phi = \phi_{\Delta S} = \frac{m}{\hbar} \int [V(x_1, t) - V(x_2, t)] dt$. The same expression describes the phase in the originally proposed electric Aharonov-Bohm experiment (1), except that in the gravitational case, the phase is proportional to the mass rather than the electric charge. We therefore identify $\phi_{\Delta S}$ with ϕ_{AB}. In our measurement, the signal of interest is the gradiometer phase shift (the difference between the phase shifts of the upper and lower interferometers due to the source mass). There is no configuration in which the interferometer trajectories are completely unperturbed. However, there is a particular approach distance ($R_x = 6$ cm) at which the gradiometer phase response to the deflections sums to zero ($\phi_{MP} = 0$) and the kinetic energy
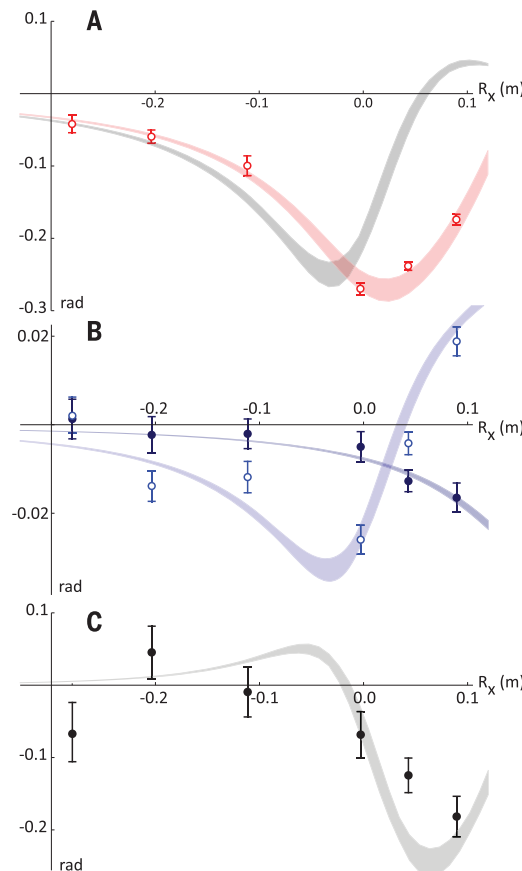

Fig. 2. Comparison of 52ℏk and 4ℏk gradiometers. (A) Phase shift induced by tungsten ring in $52\hbar k$ gradiometer as a function of R_x (red points). Theoretical predictions are based on quantum-mechanical calculation with semiclassical approximation (12) (red curve) and midpoint theorem (gray curve). The theoretical predictions are derived from ab initio models with no free parameters. Each point is the average of at least 20 shots; error bars and curve widths represent 1σ uncertainty. Curve widths are derived from uncertainty in source mass position. **(B)** Phase shifts of upper $4\hbar k$ gradiometer (light blue points) and lower $4\hbar k$ gradiometer (dark blue points) as a function of R_x, compared to theoretical predictions (light blue curve, dark blue curve). Each point is the average of at least 100 shots. **(C)** Beyond-midpoint phase shift $\phi_{\Delta S}$ of $52\hbar k$ gradiometer (black points) calculated from data in Figs. (A) and (B), compared to theoretical prediction (gray curve). $\phi_{\Delta S}$ differs significantly from zero at $R_x = 4$ cm and $R_x = 9$ cm.
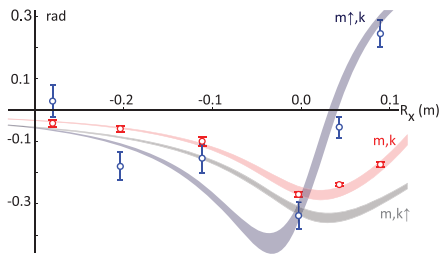
Fig. 3. Scaling properties of ϕ. Phase shift data (red points) and theoretical prediction (red curve) of 52ℏk gradiometer as a function of R_x. Near $R_x = 0$, the measured phase shift comes within 20% of the prediction for the quantum limit (large k, $\Delta x/R \gg 1$) (gray curve). Because $V \sim -GmM/R$, the red data points remain negative for $R_x > 0$. The prediction for the classical limit (large m, $\Delta x/R \ll 1$) is given by the dark blue curve. This curve changes sign near $R_x = 0$. The phase shift of the upper 4ℏk gradiometer (blue points), scaled by 52/4, agrees with the large-m curve. The large-k and large-m curves are calculated by keeping the uppermost arm trajectory and gradiometer baseline constant. Each red or blue point is the average of at least 20 or 100 shots, respectively; error bars and curve widths represent 1σ uncertainty.

contribution to the action difference also vanishes (19). Therefore, near $R_x = 6$ cm, the gradiometer phase shift is given by $\phi = \phi_{\Delta S} = \frac{m}{\hbar}\int[V(x_1,t) - V(x_2,t) - V(x_3,t) + V(x_4,t)]dt$ to within the experimental resolution, where x_3 and x_4 are the arm trajectories of the lower interferometer (20).

Together, the 4ℏk gradiometers measure the midpoint phase term ϕ_{MP} of the 52ℏk gradiometer (15). Specifically, $\phi_{MP} \approx \frac{52}{4} \cdot \frac{1}{2}(\phi_{upper} + \phi_{lower})$, where ϕ_{upper} and ϕ_{lower} are the phase shifts of the upper and lower 4ℏk gradiometers, respectively. In Fig. 2C, we subtract the measured ϕ_{MP} from ϕ to construct $\phi_{\Delta S}$ for the 52ℏk gradiometer. The signature of the Aharonov-Bohm effect is that $\phi_{\Delta S}$ is nonzero. The points at $R_x = 4$ cm (−125 ± 24 mrad) and $R_x = 9$ cm (−182 ± 28 mrad) differ significantly from zero, and the data set rejects the null hypothesis (no gravitational Aharonov-Bohm effect) with likelihood ratio 2×10^{-13}, corresponding to 7σ statistical significance. The uncertainties in Fig. 2C are limited primarily by the resolution of the 4ℏk gradiometers. As shown in Fig. 2A, the phase shift of the 52ℏk gradiometer differs from the theoretically predicted midpoint phase shift by 13σ at $R_x = 4$ cm and by 19σ at $R_x = 9$ cm. The 52ℏk data are consistent with the full phase-shift prediction (reduced $\chi^2 = 0.6$).

In a weak gravitational potential, the midpoint phase term approximately cancels with the

kinetic energy phase term [the second term in the integrand of Eq. 1; see supplementary text for details (15)]. We can therefore compute the interferometer phase shift due to the source mass as (21)

$$\phi = \frac{m}{\hbar}\int_0^{2T}[V(x_1,t) - V(x_2,t)]dt \quad (2)$$

The scaling properties of ϕ depend on the ratio $\Delta x/R$ between the wave packet separation $\Delta x \propto \hbar k/m$ and the approach distance $R = \left(R_x^2 + R_y^2\right)^{1/2}$ between the upper arm and the source mass. As m is increased at constant k and R, $\Delta x/R \to 0$, and the phase shift

$$\phi \approx \frac{m}{\hbar}\int_0^{2T} \frac{\partial V}{\partial x} \cdot \Delta x\, dt$$
$$= k\left[\int_0^T \frac{\partial V}{\partial x}t\, dt + \int_T^{2T} \frac{\partial V}{\partial x}(2T - t)dt\right] \quad (3)$$

becomes linearly proportional to k and independent of m. In this classical regime where $\Delta S = 0$, the phase shift is proportional to the gravitational acceleration $\frac{\partial V}{\partial x}$ induced by the source mass and is equal to ϕ_{MP}. By contrast, as k is increased at constant m and R, the wave packet separation increases. As $\Delta x/R \to \infty$, the phase shift $\phi \approx \frac{m}{\hbar}\int V(x_1,t)dt$ becomes independent of k and depends linearly on m. In this quantum regime, the phase shift is proportional to the gravitational potential of the source mass. Figure 3 plots the phase shifts of the 52ℏk and upper 4ℏk gradiometers as a function of R_x, comparing them to the predictions for $\Delta x/R \ll 1$ and $\Delta x/R \gg 1$. The data for the upper 4ℏk gradiometer are consistent with the classical limit.

The change in the scaling of ϕ from linear in k to linear in m is necessary to satisfy Heisenberg's error-disturbance relation, which was first introduced in Heisenberg's microscope thought experiment. This relation states that the maximum retrievable information in a quantum measurement is related to the amount of back-action it creates (22). The interferometer phase shift can be conceptualized as a measurement of the tungsten ring position R via the gravitational interaction $V \sim -GmM/R$, where M is the mass of the ring. The back-action of the atoms on the ring (i.e., the momentum recoil of the ring) is given by V and is proportional to m. Increasing the resolution of the interferometer by increasing k eventually saturates the retrievable information, as ϕ becomes insensitive to k and proportional to m, just like the back-action. Our data are consistent with Heisenberg's relation in the regime where one interaction partner is in a large quantum superposition.

The results obtained in this work are distinguished from previous gravitational mea-

surements in quantum systems by the nonlocal nature of the observed phase shift. In prior experiments with freely falling (18) and guided (23) interferometers, the wave packet separation is small enough that the gravitational field is approximately uniform at the length scale of the interferometer. The interferometer is therefore a local system in the general-relativistic sense (24). According to the equivalence principle, it is not possible to observe gravitational effects in local systems. Such experiments test the equivalence principle but do not provide any further information about the interaction of gravity with quantum particles. In addition, our experiment operates in a different regime than the Pound-Rebka experiment (25, 26), which measures the locally observable time dilation of displaced clocks due to nongravitational forces (27). The massive particles used in our interferometer measure the gravitationally induced time dilation along nonlocal trajectories, as observed in satellite experiments with classical clocks (28).

With a gravity gradient resolution of $5 \times 10^{-10}/s^2$ per shot (differential acceleration resolution $1.1 \times 10^{-11}g$ per shot, $1.4 \times 10^{-12}g$ after 70 shots), the single-source gradiometer sets a new standard for ground-based gravity gradiometry (16) and could be incorporated into proposed space-based gradiometers (29). This result is the first observation of a gravitational phase shift that is intrinsically proportional to the mass of the test particle. In addition, the phase shift depends intrinsically on Planck's constant h and Newton's gravitational constant G. Combined with a precise characterization of the source mass, this interferometer could provide an improved measurement of G (8, 30). These long-time, large-momentum-transfer interferometry techniques also enable more accurate tests of the equivalence principle (13), new searches for dark matter (31), and new types of gravitational-wave detectors (32).

REFERENCES AND NOTES

1. Y. Aharonov, D. Bohm, *Phys. Rev.* **123**, 1511–1524 (1959).
2. C. Marletto, V. Vedral, *Phys. Rev. Lett.* **125**, 040401 (2020).
3. R. G. Chambers, *Phys. Rev. Lett.* **5**, 3–5 (1960).
4. H. Batelaan, A. Tonomura, *Phys. Today* **62**, 38–43 (2009).
5. J. Gillot, S. Lepoutre, A. Gauguet, M. Büchner, J. Vigué, *Phys. Rev. Lett.* **111**, 030401 (2013).
6. J. Audretsch, C. Lämmerzahl, *J. Phys. Math. Gen.* **16**, 2457–2477 (1983).
7. M. A. Hohensee, B. Estey, P. Hamilton, A. Zeilinger, H. Müller, *Phys. Rev. Lett.* **108**, 230404 (2012).
8. G. Rosi, F. Sorrentino, L. Cacciapuoti, M. Prevedelli, G. M. Tino, *Nature* **510**, 518–521 (2014).
9. P. Wolf et al., *Class. Quantum Gravity* **28**, 145017 (2011).
10. Here and in the remainder of the text, ΔS is the gravitational contribution to the action difference. Action differences induced by photon recoil or by other electromagnetic interactions are not included.
11. C. Antoine, C. J. Bordé, *J. Opt. B Quantum Semiclassical Opt.* **5**, S199–S207 (2003).
12. C. Overstreet, P. Asenbaum, M. A. Kasevich, *Am. J. Phys.* **89**, 324–332 (2021).
13. P. Asenbaum, C. Overstreet, M. Kim, J. Curti, M. A. Kasevich, *Phys. Rev. Lett.* **125**, 191101 (2020).
14. M. Kim et al., *Opt. Lett.* **45**, 6555–6558 (2020).

15. Materials and methods are available as supplementary materials.
16. P. Asenbaum et al., Phys. Rev. Lett. **118**, 183602 (2017).
17. Although the interferometer phase is sensitive to the proper time difference between arms, the interferometer does not produce any signal at the Compton frequency mc^2/h. Our measurement does not test the universality of gravitational redshift (26).
18. R. Colella, A. W. Overhauser, S. A. Werner, Phys. Rev. Lett. **34**, 1472–1474 (1975).
19. The tungsten ring changes the velocity of the atoms on the order of 1 nm/s, which is seven orders of magnitude smaller than the recoil velocity from the interferometer beam splitters. Therefore, quadratic terms in the deflection velocity are negligible. See supplementary text for details (15).
20. Note that this condition is achieved even though the source mass is stationary in the reference frame of the laboratory.
21. P. Storey, C. Cohen-Tannoudji, J. Phys. II **4**, 1999–2027 (1994).
22. P. Busch, P. Lahti, R. F. Werner, Phys. Rev. Lett. **111**, 160405 (2013).
23. V. Xu et al., Science **366**, 745–749 (2019).
24. We use "local" to describe systems in which the length scale is small enough, and the measurement sensitivity low enough, that the effects of gravitational curvature are negligible.
25. R. V. Pound, G. A. Rebka, Phys. Rev. Lett. **4**, 337–341 (1960).
26. A. Roura, Phys. Rev. X **10**, 021014 (2020).
27. C. M. Will, Living Rev. Relativ. **17**, 4 (2014).
28. R. F. C. Vessot et al., Phys. Rev. Lett. **45**, 2081–2084 (1980).
29. O. Carraz, C. Siemes, L. Massotti, R. Haagmans, P. Silvestrin, Microgravity Sci. Technol. **26**, 139–145 (2014).
30. H. Müller, in Atom Interferometry, G. M. Tino, M. A. Kasevich, Eds. (Amsterdam, Bologna: IOS Press and SIF, 2014), pp. 389-390.
31. A. Arvanitaki, P. W. Graham, J. M. Hogan, S. Rajendran, K. Van Tilburg, Phys. Rev. D **97**, 075020 (2018).
32. M. Abe et al., Quantum Sci. Technol. **6**, 044003 (2021).
33. C. Overstreet, P. Asenbaum, J. Curti, M. Kim, M. A. Kasevich, Replication Data for: Observation of a gravitational Aharonov-Bohm effect, Harvard Dataverse, V2 (2021); https://doi.org/10.7910/DVN/O9IS7Y.

Biorepository

To support increasingly complex research, AMS Biotechnology (AMSBIO) provides a full range of blood-based products, including whole blood, serum, plasma, blood, and peripheral blood mononuclear cells. All samples are collected under institutional review board approval by certified phlebotomists and processed by standard operating procedures. Drawing upon a well-established and extensive network, AMSBIO offers custom collections according to donor demographics, sample volume, and processing method. Covering many therapeutic areas, including oncology, autoimmune, and neurodegenerative disease, samples can be delivered from patient cohorts that precisely match the unique requirements of different research studies. AMSBIO also offers a custom collection service, allowing acquisition of healthy, common, and rare-disease tissue samples according to your unique research requirements. These tissue samples can be tailored to specific donor demographic requests, tissue processing protocols, or matched sample set criteria. These tissues can be prepared in the format best suited to your research, including FFPE, fresh-frozen, or frozen optimal cutting temperature blocks. All tissue specimens are collected under ethical regulations and in accordance with all applicable local and international laws. Using state-of-the-art equipment, AMSBIO can identify mutation-specific single nucleotide polymorphisms, fusions, and rearrangements; perform NGS analyses on your tumor tissues of interest; and identify overexpression.

AMS Biotechnology
For info: +1-617-945-5033
www.amsbio.com/products/biorepository

Electronic Pipettes

INTEGRA Biosciences provides an extensive range of multichannel and electronic pipettes as well as robotic-assisted platforms to increase the throughput of your cell culture workflows. Electronic pipettes and automated systems have been proven to offer significant advantages over manual processes, providing reproducible results with fewer handling errors, while eliminating the painstaking work of repetitive liquid handling. INTEGRA has a variety of solutions to meet your liquid-handling needs, whether you're looking to upscale to a multichannel pipette or increase your throughput with automation. The company's range of sophisticated 4- to 384-channel multichannel handheld pipettes allows you to simultaneously pipette entire rows, columns, or even plates for greater throughput. With INTEGRA electronic pipettes, you can program and store pipetting protocols and mixing routines for use across various workflows.

INTEGRA Biosciences
For info: +1-603-578-5800
www.integra-biosciences.com

Single-Cell Dispenser

By automating labor-intensive and time-consuming steps, F.SIGHT 2.0 significantly streamlines cell-line development, cell and gene therapy, and single-cell omics workflows. F.SIGHT 2.0 is a fast single-cell dispenser that isolates both unlabeled and green fluorescent cells. With our patented, highly efficient, and gentle single-cell microfluidic technology, single cells are dispensed into a 96-well plate in less than 2 min or a 384-well-plate in less than 7 min. The innovative dual-camera system enables you to simultaneously capture brightfield and fluorescence information at full resolution. You can control both cameras independently and leverage the powerful, built-in, and free adjustable blue laser. Combined, their high dynamic range visualizes even the weakest fluorophores. Sorting individual cells into microwell plates can be time-consuming and difficult; especially when the cells are limited or fragile, the experiment requires high cell viability, or small volumes of lysis buffer are used for downstream reaction miniaturization. F. SIGHT 2.0 allows precise, reliable dispensing even into PCR plates with <1 µL of lysis buffer, and can handle low cell numbers in only 5 µL of suspension. Its intuitive interface and handling requires minimal user training.

Cytena
For info: +1-833-235-5465
www.cytena.com

Syringeless Filters

Mini-UniPrep G2 is an all-in-one syringeless filter with a borosilicate glass chamber. It replaces syringe filters for autosampler-supported UHPLC/HPLC sample preparation with easier handling and fewer consumable components. The unit features an integrated autosampler vial, plunger with attached filter membrane, and septum/cap. Borosilicate glass vials eliminate plastic-based leachables that can originate from polypropylene vials. Mini-UniPrep G2 holds up to 0.5 mL of unfiltered sample that can then be loaded directly into an autosampler. It is easy to see when a sample has been filtered. These filters are featured in a wide range of membranes with 0.2- and 0.45-µm pore sizes to meet specific requirements, and are also available in amber glass for light-sensitive samples, and in polypropylene format designed for autosamplers.

Cytiva Life Sciences
For info: +1-800-526-3593
www.cytivalifesciences.com

Custom Beamsplitters

Optical Surfaces is a leading supplier of single to original equipment manufacturer (OEM) quantities of high-precision, custom beamsplitters to a growing number of spectroscopy, interferometry, and imaging instrument manufacturers worldwide. The optimal beamsplitter can take on many forms, including cubes, plates, hexagons, pentagons, polarizing, nonpolarizing, or wedge designs in a variety of reflection/transmission ratios and substrates, depending upon your application. Drawing on over 50 years of experience, we can advise you on key parameters to consider, including wavelength range, physical size, polarization, and—if used in an interferometer—pathlength and phase-matching requirements. Our beamsplitters offer outstanding flatness and parallel specifications, and are manufactured to a surface accuracy of lambda/20 peak to valley (P-V) at 633 nm, a surface quality of 10/5, and microroughness of less than 1.2-nm root mean square (rms). Supplied mounted or unmounted, they can be enhanced with an extensive range of coatings designed to operate at your required angle of incidence.

Optical Surfaces
For info: +44-(0)-208-668-6126
www.optisurf.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.
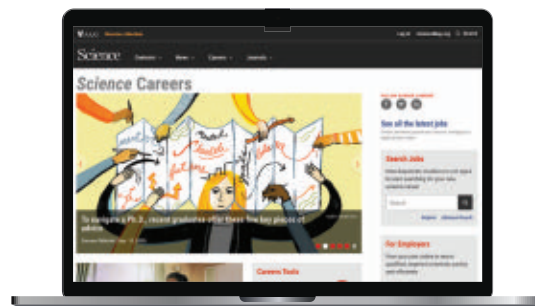
No need to hide

Growing up on a farm in northeastern Brazil, I learned from my parents about the large bees that frequented our passion fruit crop—*mangangava*, we called them. But in college, when I gave a presentation about those bees, I didn't mention their familiar name; instead I used formal taxonomic terms when describing their adaptations for pollinating. I was afraid of being judged for expressing my personal knowledge in a science class. But when the instructor pressed me to tell the class whether I'd seen the bees myself and what I called them, my views—and my whole career trajectory—began to change. I realized my background was an asset, and that my calling was to help educate others like me.

I was the first and only person in my family to graduate from high school and pursue higher education. My mother only went to elementary school and my father had no schooling at all. They worked on a small farm inherited from my grandparents, growing potatoes, cassava, oranges, passion fruit, and other crops. I helped out when I was not in school.

My parents always had a close relationship with nature, something they learned from my grandparents and passed down to me. For example, they knew the exact time of year to plant or pick certain fruits, when there would be a long rainy or dry season, which insects frequented the plants, and which ones could be dangerous when threatened. This was the main reason I became interested in biology.

In college, I joined a series of labs to gain research experience, started to publish papers, and attended scientific meetings. But my interest in pursuing a career as a scientist began to wane after I realized my research was not helping poor people like my parents. I also felt out of place in most classes. As a gay, Black student from a rural community, I did not see myself in my professors, most of whom were straight and white and spoke with a different accent.

The "top-down" instruction heightened my sense of alienation: The professors would talk and we, the students, would listen and take notes. Even if I studied for several hours per day, I fell behind peers who came from private schools. The feeling of not belonging circled in my head. I sometimes thought about giving up.

My instructor's request that I share my knowledge of the *mangangava* bee changed my perspective. She saw that tapping into my personal experiences would help me grasp

"The best way to nurture new scientists is to allow students to be themselves."

the scientific concepts. No other professor had shown that kind of sensitivity, which helped me realize the value of enlisting students' own knowledge and perspectives.

I decided to switch my major to focus on science education. As part of that program, I met a professor who came from a town close to where I grew up. I asked whether I could intern with her and was pleased to discover that she, too, was interested in traditional knowledge. Under her guidance, I created a book that taught elementary school students about the importance of insects, telling a story that wove together scientific concepts and traditional knowledge.

Both instructors made it clear they valued what I'd learned from my family and from my personal experiences on the farm. They also had identities that resonated with me, as one was transgender and the other was Black. They helped me see I did have a place in higher education and that I didn't need to hide who I truly am.

I'm now a master's student in education, working on a project that explores the intersection between science and traditional knowledge using a form of poetry that's popular in northeastern Brazil. I don't think I'd be here today if I hadn't crossed paths with those two instructors. We need more educators and scientists like them—people who can serve as role models and know that the best way to nurture new scientists is to allow students to be themselves. ∎

Anderson Eduardo-Santos is a master's student at the University of São Paulo, Ribeirão Preto. He also works as a science teacher at a public elementary school. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Where Science Gets Social.
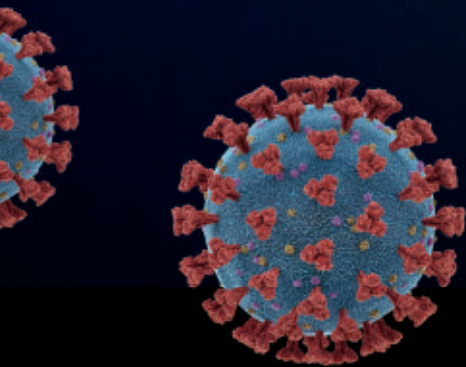
AAAS.ORG/COMMUNITY

AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY
AAAS

Accelerate your contribution to the fight against COVID-19

Science is how we end COVID-19. The COVID-19 High Performance Computing Consortium is still active and wants to help researchers worldwide accelerate the pace of scientific discovery in the fight against the virus.

Apply today to leverage the computing resources of our partner institutions, at no cost, and help us end this pandemic.

covid19-hpc-consortium.org

In March 2020, we brought together more than 40 universities, government agencies, and companies—many of whom are usually rivals—with the goal of utilizing their compute resources and experience to help turn the tide on COVID-19.

The COVID-19 HPC Consortium has supported over 100 projects leading to real discoveries, from understanding how the virus interacts with human receptor proteins to new molecules and proteins able to combat the virus.

Our work is not done. We're continuing to support projects that can make a difference in the find to end this pandemic.

We're looking for projects in areas such as:

- Understanding and modeling patient response to the virus using large clinical datasets

- Learning and validating vaccine response models from multiple clinical trials

- Evaluating combination therapies using repurposed molecules

- Mutation understanding and mitigation methods

- Epidemiological models driven by large multi-modal datasets