The race for room-temperature superconductors heats up p. 954

Lifelong effects of prenatal infections pp. 967 & 982

Thin, conductive, stretchy membranes p. 1022

Science

$15
27 AUGUST 2021
sciencemag.org

AAAS



A WINNING SOLUTION

Artificial intelligence reveals RNA structure pp. 964 & 1047



Presented by BII & Science

BII BioInnovation Institute Science

Apply before November 1, 2021
www.bii.dk/scienceprize

CONTENTS

27 AUGUST 2021 • VOLUME 373 • ISSUE 6558



NEWS

IN BRIEF

946 News at a glance

IN DEPTH

948 **Afghan scholars despair after Taliban's takeover**
Many scramble to leave the country in fear of return to harsh "antiscience" rule
By R. Stone

949 **Unethical? Unnecessary? The booster debate intensifies**
As United States reveals its plan to offer an extra dose of COVID-19 vaccine, equity and scientific questions abound *By G. Vogel*

950 **Honesty study was based on fabricated data**
Made-up data set raises questions about behavioral scientist Dan Ariely
By C. O'Grady

952 **Scaled down, martian model habitat rises again in desert**
Leaders have big dreams for greenhouse at Biosphere 2 *By M. Price*

953 **Looking to the Rockies for clues to water woes**
New effort to understand precipitation and runoff could improve Colorado River forecasts *By E. Stokstad*

FEATURES

954 **Feeling the pressure**
Can room-temperature superconductors transform from a tantalizing glimmer to practical materials?
By R. F. Service

INSIGHTS

POLICY FORUM

958 **Allow "nonuse rights" to conserve natural resources**
"Use-it-or-lose-it" requirements should be reconsidered *By B. Leonard* et al.

PERSPECTIVES

962 **Driving multiphase superconductivity**
Breaking local symmetry opens up a clear transition between superconducting states
By A. Pourret and G. Knebel
RESEARCH ARTICLE p. 1012



963 **Pushing low thermal conductivity to the limit**
Weak bonding between the distorted layer of $Bi_4O_4SeCl_2$ helps limit its heat transport
By S. E. Kim and D. G. Cahill
REPORT p. 1017

964 **Piercing the fog of the RNA structure-ome**
Machine learning is poised to transform RNA structure and function discovery *By K. M. Weeks*
REPORT p. 1047

966 **How microbiota improve immunotherapy**
Ligands derived from the gut microbiota enhance cancer immunotherapy
By E. Ansaldo and Y. Belkaid
REPORT p. 1040

967 **Immune imprinting in utero**
Mild maternal infection causes tissue-specific epigenetic imprinting in utero *By M. Amir and M. Y. Zeng*
RESEARCH ARTICLE p. 982

968 **The animal origin of SARS-CoV-2**
Trading of animals susceptible to bat coronaviruses is the likely cause of the COVID-19 pandemic *By S. Lytras*

BOOKS *ET AL.*

971 **The spectrum of happiness**
A pair of exhibitions challenge visitors to embrace a more encompassing view of contentment *By V. Skvortsova*

PHOTOS: (TOP TO BOTTOM) BOB WICK/BUREAU OF LAND MANAGEMENT; J. ADAM FENSTER/UNIVERSITY OF ROCHESTER

972 The unsung players of epidemiology
A new history probes often-overlooked contributions to the study of infectious disease *By S. Seth*

LETTERS

973 Feral equids' varied effects on ecosystems
By E. S. Rubin et al.

973 Response
By Erick J. Lundgren et al.

974 Academic bullying: How to be an ally
By M. Mahmoudi

RESEARCH

IN BRIEF

977 From *Science* and other journals

REVIEW

980 Structural biology
Advances and challenges in time-resolved macromolecular crystallography
G. Brändén and R. Neutze
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABA0954

981 Coronavirus
Airborne transmission of respiratory viruses
C. C. Wang et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD9149

RESEARCH ARTICLES

982 Immunology
Prenatal maternal infection promotes tissue-specific immunity and inflammation in offspring *A. I. Lim* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABF3002
PERSPECTIVE p. 967

983 Neurodevelopment
Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells *I. Sarropoulos* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABG4696

984 Molecular biology
RNA editing restricts hyperactive ciliary kinases *D. Li* et al.

991 Coronavirus
Chimeric spike mRNA vaccines protect against Sarbecovirus challenge in mice
D. R. Martinez et al.

998 Quality control
Identification of a quality-control factor that monitors failures during proteasome assembly *E. Zavodszky* et al.

1004 Organic chemistry
Photomediated ring contraction of saturated heterocycles *J. Jurczyk* et al.

1012 Superconductivity
Field-induced transition within the superconducting state of $CeRh_2As_2$
S. Khim et al.
PERSPECTIVE p. 962

REPORTS

1017 Thermal transport
Low thermal conductivity in a modular inorganic material with bonding anisotropy and mismatch *Q. D. Gibson* et al.
PERSPECTIVE p. 963

1022 Applied physics
Highly conductive and elastic nanomembrane for skin electronics *D. Jung* et al.

1027 Paleoecology
A positive relationship between functional redundancy and temperature in Cenozoic marine ecosystems *T. M. Womack* et al.

1030 Human genetics
Population sequencing data reveal a compendium of mutational processes in the human germ line *V. B. Seplyarskiy* et al.

1035 Paleoclimate
Insolation triggered abrupt weakening of Atlantic circulation at the end of interglacials
Q. Z. Yin et al.

1040 Microbiome
Enterococcus peptidoglycan remodeling promotes checkpoint inhibitor cancer immunotherapy *M. E. Griffin* et al.
PERSPECTIVE P. 966

1047 RNA
Geometric deep learning of RNA structure
R. J. L. Townshend et al.
PERSPECTIVE p. 964; PODCAST



DEPARTMENTS

945 Editorial
Colleges must require vaccination
By Michael A. McRobbie

1054 Working Life
Waking up to my sleep disorder
By Ashley M. Bourke

ON THE COVER

An artificial intelligence algorithm selects an RNA molecule's three-dimensional shape out of a sea of incorrect shapes. Computational prediction of the structures into which RNAs fold is particularly important—and particularly difficult—because so few structures are known. The success of machine learning despite this lack of data opens doors to understanding and designing diverse molecules, including medicines. See pages 964 and 1047. *Illustration: C. Bickel/Science; Data: Andrew M. Watkins and RNA-Puzzles consortium/rnapuzzles.org*





AAAS News & Notes 975
Science Careers 1052

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2021 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.sciencemag.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

CREDITS: (PHOTO) MICHAEL ALFUSO; (ILLUSTRATION) ROBERT NEUBECKER

EDITORIAL

Colleges must require vaccination



Michael A. McRobbie is president emeritus and university professor at Indiana University, IN, USA. mcrobbie@iu.edu

This week, the US Food and Drug Administration (FDA) granted full approval to the Pfizer-BioNTech coronavirus vaccine for individuals 16 years and older. This milestone decision, the first full approval of a vaccine for COVID-19, will almost certainly clear a path for businesses, hospitals, and government agencies that have not already done so to adopt vaccine mandates for their employees. For colleges and universities that have been on the fence about requiring the vaccine, the FDA's decision may be especially welcome news.

As of earlier this week, 753 campuses require the vaccination. According to a map from the Chronicle of Higher Education, most of these schools are in "blue" states. There is no doubt that COVID-19 vaccination has become politicized. Now, with the full FDA approval, there is even less reason for the political hue of a state to deter universities, as citadels of science and reason, from making every attempt to implement vaccine mandates.

Indiana University, a public institution where I stepped down as president at the end of June, announced a COVID-19 vaccine mandate for all students, faculty, and staff in May, just as the Delta variant was starting to surge in the United States. Indiana is a "red" state, and this decision triggered one of the first legal disputes in the nation over the college vaccine mandate.

The university's public health experts were emphatic that vaccination would be the only way to ensure a return to mostly normal operations and a more typical university experience this fall. And the university's leadership viewed the mandate, which includes appropriate exemptions, as a continuation of its science- and public health–driven approach to manage the pandemic on all of its campuses throughout Indiana.

Opposition to this decision was expected, and just days after announcing the mandate, 35 state senators sent me a strongly worded letter urging the university to reconsider and rescind the mandate.

"We find ourselves on the cusp of victory over an enemy we have come to know as the novel coronavirus…," the senators wrote. "After 14 months of fighting and enduring the COVID-19 war, our state is finally returning to the path of normalcy. Regrettably, decision-makers at Indiana University have veered away from that path."

In late June, eight students filed a lawsuit against Indiana University, alleging that the vaccine mandate violated their constitutional rights by forcing them to receive unwanted medical treatment. The following month, a federal judge decisively rejected this argument on the grounds that students, in fact, do have options—they can get vaccinated, apply for an exemption, or choose to attend another school.

Earlier this month, an appeals court unanimously upheld the judge's decision, and on 12 August, Supreme Court Justice Amy Coney Barrett denied a bid to block Indiana University's mandate. This third ruling, now from the nation's highest court, signals that similar vaccine requirements are very likely to pass legal muster.

Public data, as well as the consensus of medical and scientific opinion, indicate that the battle against COVID-19 is far from over. Vaccination rates in many areas of the United States simply are not high enough to prevent enough COVID-19 infections and transmission. In Indiana, COVID-19 case rates are the highest they have been since May, and vaccination rates languish. The highly contagious Delta variant is stoking a deadly new surge of cases, hospitalizations, and deaths, especially in places with low vaccination rates.

> **"The Delta variant has been bad. The next variant could be worse."**

As the nation's top public health experts have repeatedly expressed, vaccine mandates, which continue to receive high public support, represent the best path to increase vaccination numbers to the levels necessary to defeat the virus and avoid the risk of continuing mutations and variants. The Delta variant has been bad. The next variant could be worse.

Fortunately, more universities are deciding that the situation is serious enough to require vaccine mandates. And the American College Health Association released a statement this month, which was signed by more than two dozen higher education organizations, condemning state-level restrictions that bar colleges from requiring the COVID-19 vaccine or adopting other public health measures, such as masking.

For all states—red, blue, or purple—the message is clear. Political differences must be put aside in the interest of public health and safety at the state's colleges, universities, and the communities that they serve through their vital education and research missions.

– Michael A. McRobbie

PHOTO: INDIANA UNIVERSITY

10.1126/science.abl9102

NEWS

> “The moment you’ve been waiting for is here.”
>
> **U.S. President Joe Biden**, addressing the nation—and vaccine-hesitant Americans—after the Food and Drug Administration granted full approval to the Pfizer-BioNTech COVID-19 vaccine on 23 August.

IN BRIEF

Edited by **Catherine Matacic**



Olympic gold medalist Caster Semenya at a 2019 race in France.

PHOTO: ANTHONY DIBON/ICON SPORT VIA GETTY IMAGES

SPORTS SCIENCE

Testosterone study doesn’t go the distance

A 2017 study that suggested some female runners with naturally higher testosterone levels had a competitive advantage was corrected last week, raising doubts about its use to ban South Africa’s Caster Semenya and others from competing in events including the Tokyo Olympics. The *British Journal of Sports Medicine* study, which analyzed testosterone levels in elite athletes in the 2011 and 2013 World Athletics Championships, found that women in the top tertile for the hormone performed nearly 3% better than those in the bottom third in the 400-meter event and almost 2% better in the 800-meter event (*Science*, 27 July 2018, p. 322). That prompted a 2018 rule by the International Amateur Athletic Federation (now called World Athletics) that prevents female athletes with naturally high testosterone levels from competing in races from 400 to 1500 meters, unless they take testosterone-reducing drugs. In the 17 August correction, the paper’s authors acknowledged they had found no causal link between high levels of the hormone and enhanced performance. Semenya is challenging her ban, and track and field’s governing body hasn’t said whether it will revise its rule. Other affected athletes have switched to different races, or retired.

CDC sets up outbreak center

PUBLIC HEALTH | In an effort to build on lessons learned during the COVID-19 pandemic and improve public health decision-making, the U.S. Centers for Disease Control and Prevention (CDC) last week announced the launch of a new center to forecast outbreaks of infectious diseases and devise more effective strategies when they occur. CDC has recruited top researchers—including Harvard University epidemiologist Marc Lipsitch as director of science—to run the Center for Forecasting and Outbreak Analytics. The facility, billed as “a hub for innovation and research on disease modeling,” will also expand data-sharing networks between researchers and first-line responders. It has up to $500 million in funding from the American Rescue Plan and will begin operations in 2022.

U.S. public backs evolution

PUBLIC OPINION | A solid majority of U.S. adults now accepts evolution, a shift that researchers attribute to a growth

in scientific literacy and a decline in the politicization of the issue. University of Michigan, Ann Arbor, political scientist Jon Miller has been surveying Americans about evolution since 1985 and, in a paper out last week, he documents a steady rise in support since 2007 for the statement "human beings, as we know them today, developed from earlier species of animals." For the preceding 2 decades, a series of surveys had revealed a roughly 40% to 40% split, with 20% undecided. But by 2020, that had shifted to 53% in favor and 36% opposed, with 11% agnostic. Miller attributes the change to more people attending college, where most students are required to take at least one science course, as well as to the waning power of evolution as a wedge issue. "There are very few politicians left who play the chimpanzee card," he says. However, the United States still trails most of the industrialized world, where public acceptance of evolution can top 80%.

HHMI fires MIT biologist

#METOO | The Howard Hughes Medical Institute (HHMI) on 20 August fired prominent biologist David Sabatini after an investigation found that he violated sexual harassment and other workplace policies. Sabatini also resigned from the Whitehead Institute, the nonprofit research organization where his large HHMI-supported lab was located. Whitehead had hired an outside law firm to conduct the probe. Sabatini, who remains a professor at the Massachusetts Institute of Technology (MIT), co-discovered mTOR, a protein that is a central regulator of cell growth and aging in mammals. MIT said in a statement that its senior administration is reviewing the Whitehead Institute's findings "and determining next steps in response … up to and including revocation of tenure proceedings."

Preprint ban under review

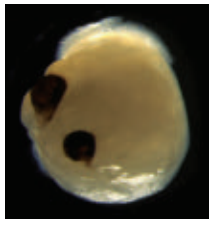
FUNDING | More than 600 researchers have signed an open letter demanding that the Australian Research Council (ARC) scrap a new policy that forbids grant applicants from mentioning preprints in their funding proposals. Applicants whose proposals had been automatically rejected took to social media last week to criticize the policy, many noting that they had not been aware of the rule, first introduced in September 2020. Proposals were rejected even if they cited preprints authored by other scientists. Some researchers also said job offers had been rescinded because those offers were dependent on securing ARC grants. An ARC spokesperson has told *Science* that the agency is reviewing the policy following feedback from the research community.



IN FOCUS After 55 years of staring at the Sun, the McMath-Pierce Solar Telescope and its 10-story building (above) will soon be transformed, with $4.5 million from the National Science Foundation, into an astronomy outreach center. Visitors to Kitt Peak National Observatory in Arizona will see three solar telescopes and interactive exhibits. And U.S. solar astronomers already have a replacement: the Daniel K. Inouye Solar Telescope, the world's largest, which opened in Hawaii in 2020.

Minibrains grow crude 'eyes'

DEVELOPMENTAL BIOLOGY | In an effort to mimic early eye development, researchers have created tiny orbs of human brain tissue with rudimentary pairs of light-sensitive, eyelike structures in a lab dish. Past studies have re-created a key structure in the eye—the light-receiving retina—by "reprogramming" mature human cells into stem cells and then guiding their development in a specially designed nutrient broth.



Two cup-shaped protrusions form rudimentary "eyes" in lab-grown brain tissue.

But in a developing embryo, these early eye structures form from tissue at the front of the brain. Hoping to more faithfully re-create this process, researchers first turned reprogrammed human stem cells into the precursors of neurons typically found in the front of the brain. These cells spontaneously formed clumps of forebrain tissue, each with two cup-shaped protrusions that became electrically active in response to light. The organoids, described last week in *Cell Stem Cell*, could help scientists better understand the origins of retinal disease.

Preventive combo curbs malaria

INFECTIOUS DISEASE | Malaria vaccines and antimalarial pills by themselves often fail to prevent the fevers, hospitalization, and death caused by the disease. Now, a randomized, controlled study involving nearly 6000 children in Mali and Burkina Faso shows participants given both interventions during malaria season fared better than those who only received one or the other. Reported this week in *The New England Journal of Medicine*, the study enrolled children between 5 and 17 months old and followed them for 3 years. Those given both the pills and a vaccine known as RTS,S had about 60% protection against symptomatic disease compared with those who received either intervention alone; the combination's efficacy was more than 70% for hospitalization or death.

IMAGES: (TOP TO BOTTOM) KPNO/NOIRLAB/NSF/AURA/D. SALMAN; GABRIEL *ET AL*., *CELL STEM CELL* 28, 1–18, 7 OCTOBER 2021 © 2021 ELSEVIER INC.



IN DEPTH

A May 2019 graduation ceremony at the American University of Afghanistan. Three years earlier, an attack on the university by suspected Taliban fighters killed 13 people.

PHOTO: SCOTT PETERSON/GETTY IMAGES

ASIA

Afghan scholars despair after Taliban's takeover

Many scramble to leave the country in fear of return to harsh "antiscience" rule

By **Richard Stone**

Twice the Taliban tried to kill Khyber Mashal*. Its first attempt was in 2009, when the Afghan scientist was working on a development project for the U.S. Agency for International Development in Gardez, in southeastern Afghanistan. Taliban fighters planted a bomb below his office, says Mashal, who was away on a short trip to Germany at the time. Five colleagues died in the blast. Then in July 2019, when Mashal was working for Afghanistan's Ministry of Education, a suicide bomber staggered in front of his car in Kabul. "He seemed intoxicated," he says. A quick-thinking police officer tased the man and removed his explosives-laden vest.

Why is the Taliban so eager to take him out? "Because they're antiscience," Mashal says. "Educated people are targeted because we have transformed the country." His past affiliation with a U.S. organization added to the jeopardy. Mashal left Afghanistan with his wife in December 2020 for a yearlong fellowship at a German university. Now, after the Taliban's lightning-fast takeover of the country, many other scientists are trying to join the exodus.

*The name Khyber Mashal is a pseudonym.

Afghanistan had come a long way since the Taliban last ruled from 1996 to 2001 under a harsh interpretation of Sharia law in which it deprived women of civil liberties and summarily executed intellectuals and others opposed to its ideology. After its ouster, Afghanistan's higher education institutions burgeoned from a handful to more than 100, and women entered the workforce en masse.

Taliban leaders insist they have moderated their views, but few Afghans are willing to take those reassurances at face value. As recently as 2016, an attack by suspected Taliban fighters on the American University of Afghanistan killed 13 people and wounded more than 50.

The gains women have made in Afghan society "will fade and be eliminated," predicts an engineer at Avicenna University, a private institute opened in Kabul in 2010, who asked to remain anonymous because she says her life is in danger. "The future is very dark" for scholars who remain in Afghanistan, says Mohammad Assem Mayar, a water management expert at Kabul Polytechnic University who worked with scientists at the University of California, Irvine, and the U.S. Geological Survey to model flood risks.

Mayar recently found succor at the University of Stuttgart, but colleagues marooned in Afghanistan dread the coming days. The Avicenna engineer, who has also collaborated with U.S. researchers, says she and her family fled their apartment in Kabul earlier this week. "The Taliban were going door to door looking for us," she says. She and her family applied for U.S. visas 6 years ago but are still waiting for a decision. Now, she's counting on colleagues in the United States to pull strings on their behalf. She has "no hope" of surviving in Afghanistan.

As *Science* went to press, European and U.S. officials were scrambling to get her and hundreds of other Afghan scholars and their families out of Kabul. Last week, several members of the Afghan robotics team—young women who gained fame for their ingenuity in international competitions—managed to leave on flights to Qatar. But reaching Kabul's airport means running a gauntlet of Taliban fighters and passing several checkpoints. Scientists stranded outside the capital, including a female scientist holed up in a basement in the western city of Herat, say it's too dangerous to travel to Kabul now.

In an 18 August letter with more than 3400 signatories, Robert Quinn, executive director of the Scholars at Risk (SAR) Network, urged U.S. Secretary of State Antony Blinken to relax visa requirements for Afghans and continue evacuation flights until all "scholars, students, and civil society actors" are out.

SAR has received hundreds of pleas for assistance over the past several days, Quinn says.

Things could turn very bleak very fast for scholars left behind—even those not in the Taliban's crosshairs. Mayar anticipates that a cash-strapped Taliban regime is unlikely to pay salaries to university faculty and staff, as happened during its previous rule. "There's also a high potential that academic facilities will be looted," says Alex Dehgan, who as country director for the Wildlife Conservation Society from 2006 to 2008 helped establish Afghanistan's first national park, Band-e-Amir.

Academic life is inimical to the Taliban's ideology, Mashal says. "Hardly anybody in the Taliban leadership is educated," he says. Rank-and-file fighters are mostly "brainwashed kids coming out of the madrassahs. They're trained to only think about two things: heaven and hell." In a sign that the Taliban intends to assert control over universities, it installed its own rector at Paktia University in Gardez. "Their idea is to handicap these institutions. Push them back to the first century," Mashal says.

Many U.S. institutions are trying to protect former collaborators by purging their websites and social media accounts of any mention of past cooperation. And they are coordinating with Biden administration officials and Congress on how to steer scholars to safe harbors.

Michigan State University's Grain Research and Innovation Project, for example, has trained researchers in Kabul in recent years and placed 33 students—including 12 women—in graduate programs at Kabul University and two Indian universities. Now, the program is trying to rescue the researchers it nurtured. "We want to find a good home for them where they can practice their science, where they can raise their families, where they can be safe and secure," says project director Kurt Richter. From his haven in Germany, Mashal has arranged a visa for an Afghan student to come to his university for doctoral studies.

With his own fellowship ending in November, Mashal has applied to SAR for his next lifeline. He has no intention of returning to Afghanistan while the Taliban is in power: "I don't want to die." But he agonizes over the perils his friends and remaining family members face back home. As Kabul was about to fall, he developed "extreme anxiety" and now has trouble sleeping. "I try to console my colleagues. I try to console my family. And I try to console myself," he says. "But it's so painful to see the devastation. The loss of everything we risked our lives for." ■

Richard Stone is senior science editor at the Howard Hughes Medical Institute's Tangled Bank Studios.

COVID-19

Unethical? Unnecessary? The booster debate intensifies

As United States reveals its plan to offer an extra dose of COVID-19 vaccine, equity and scientific questions abound

By **Gretchen Vogel**

As the extraordinarily infectious Delta variant of SARS-CoV-2 continues to spread around the world, vaccines' powers are showing their limits. Although they are still extremely effective at preventing severe COVID-19, the tantalizing hope that the shots could block almost all infections—and squelch transmission—has evaporated. That has upended return to office and school plans, threatened economic recoveries, and spurred fresh political rows over mask and vaccination mandates.

Now, amid hints that vaccine-induced immunity is waning, policymakers and scientists are debating whether widespread booster shots could help—or whether getting shots into the arms of the unvaccinated should remain the top priority. And many people wonder whether one booster will suffice or periodic COVID-19 vaccination will become the new normal, as it is for influenza.

> **"You are reducing supply for those who need it more."**
>
> **Bruce Aylward**, World Health Organization

On the latter question, some scientists say experience with other vaccines suggests a single well-timed booster may provide long-lasting immunity. But others contend the booster rush is premature given scant data on their effectiveness and best timing. "We don't understand who is going to need a booster, how long after their last dose, or which vaccine combination works best," says physician-epidemiologist Bruce Aylward, a senior adviser at the World Health Organization (WHO). "You need to understand all that before you decide how boosters should be used."

That hasn't stopped booster rollouts in countries like Israel, which is seeing a Delta-fueled surge of COVID-19 cases, hospitalizations, and deaths, despite one of the earliest and fastest vaccine campaigns in the world. More than 60% of the Israeli population has received two doses of Pfizer's messenger RNA (mRNA) vaccine, but on 30 July, Israel began to offer a third dose of the vaccine to anyone 60 and older—the first country to do so. On 20 August, it said everyone 40 and older should get one.

WHO and other organizations have warned strongly against such broad booster rollouts, mainly because many high-risk people worldwide have not even received a first vaccine dose. Giving boosters now "is unfair to say the least, potentially … even criminal," says Tulio de Oliveira, a computational biologist at the University of KwaZulu-Natal, Durban, who has used sequencing to track the pandemic's spread in Africa.

The United States and several other countries have started to give boosters to people with weakened immune responses or at high risk of developing serious COVID-19. Some regions are also recommending boosters for health care workers, as even mild breakthrough cases can hobble the health system if staff have to isolate. Giving a third dose to those smaller groups is less controversial; the expected benefits are somewhat clearer, and the numbers of COVID-19 vaccine doses required only have a marginal impact on global supplies, Aylward says.

But other wealthy countries may well follow Israel's lead and make boosters available to nearly everyone. On 18 August, just days before the U.S. Food and Drug Administration gave full approval to Pfizer's COVID-19 vaccine, the Biden administration said it would provide boosters more broadly in September, although that would first require approval by FDA and by a U.S. panel that recommends which specific populations should receive immunizations.

Many vaccine experts argue there isn't enough evidence that boosters are needed or will truly help control the pandemic, especially because multiple studies show existing vaccine regimens are holding strong against severe disease. A large study of patient health records in New York state, published on 18 August in the *Morbidity and Mortality Weekly Report*, found that vaccine efficacy against all SARS-CoV-2 infections dropped from 91.7% to 79.8% between May



Israel has started to administer COVID-19 booster shots to people ages 40 and older.

and July, as Delta took over in the region. But protection against hospitalization for COVID-19 stayed close to 95%. Data from the Israeli Ministry of Health suggests protection against severe disease is still nearly 92% for people 50 and younger and 85% for those older than 50.

That suggests boosters are the wrong way to use the world's still limited vaccine supply, says Aylward, who helps coordinate the COVID-19 Vaccines Global Access Facility, which distributes doses to low- and middle-income countries. If everyone in high-income countries received boosters, that would use up 1 billion doses, Aylward estimates. "You're dealing with a finite, zero-sum resource," he says. "You are reducing supply for those who need it more."

Even preparing for possible boosters disrupts the supply and distribution system, he says, as countries stock up on extra doses. "The U.K. has 66 million people and recently bought another 110 million doses—and they already have 80% of the [eligible] population vaccinated," de Oliveira says, "while at the moment Africa is at less than 3%."

Others don't deny that boosting in rich countries might disadvantage the rest of the world, but say that, scientifically speaking, a third dose is likely to help shore up the immune system. "[Vaccine] efficacy drops with Delta. That is indisputable," especially for mild disease, says Leif Erik Sander, an infectious disease expert at the Charité University Hospital in Berlin.

Other vaccines require three shots to confer long-lasting protection, including those for hepatitis B and tick-borne encephalitis. And some scientists have suggested the current COVID-19 vaccine dose spacing—just 3 to 4 weeks between mRNA doses, for example—was chosen not to provide long-lasting immunity, but to speed clinical testing. A boost months later may be ideal, they say, and doesn't necessarily mean yearly updates will be needed. Although SARS-CoV-2 is evolving in dangerous ways, it does not seem to undergo the same type of genetic shuffling flu viruses do in other animals.

"A third dose is a good idea," says Akiko Iwasaki, an immunologist at Yale University. Although at-risk groups should receive them first, she adds, "if there are enough doses, I think the general public will benefit," both because they are likely to help reduce SARS-CoV-2's spread and because even mild COVID-19 can lead to long-term complications.

In Israel, there are signs that the third shots are reducing infections in older people. "The very preliminary data look like it is having an effect," says Barak Raveh, who studies dynamic systems at the Hebrew University of Jerusalem. His analysis of the Health Ministry's data suggests protection against SARS-CoV-2 infection for vaccinated people has climbed in recent weeks, likely driven by boosters; 80% of people in their 70s have already received a third dose.

There's no telling how long that fortified immune shield will last. Multiple trials are underway to more precisely measure the effects of boosters—including mixing and matching different types of COVID-19 vaccines—but they will likely take several more months. As with so much in the pandemic, policymakers will have to make far-reaching decisions long before clear answers are in. ■

With reporting by Meredith Wadman.

SCIENTIFIC INTEGRITY

Honesty study was based on fabricated data

Made-up data set raises questions about behavioral scientist Dan Ariely

By **Cathleen O'Grady**

Dan Ariely is a behavioral science superstar. His research on honesty, cheating, and irrationality is "extremely clever and extremely intuitive," says behavioral scientist Eugen Dimant of the University of Pennsylvania—and it has had a huge impact on both the field and government policies. Ariely, who founded the Center for Advanced Hindsight at Duke University, has also written three *New York Times* bestsellers and is a TED Talks regular.

But some researchers are calling Ariely's large body of work into question after a 17 August blog post revealed that fabricated data underlie part of a high-profile 2012 paper about dishonesty that he co-wrote. None of the five study authors disputes that fabrication occurred, but Ariely's colleagues have washed their hands of responsibility for it. Ariely acknowledges that only he had handled the earliest known version of the data file, which contained the fabrications.

Ariely emphatically denies making up the data, however, and says he quickly brought the matter to the attention of Duke's Office of Scientific Integrity. (The university declined to say whether it is investigating Ariely.) The data were collected by an insurance company, Ariely says, but he no longer has records of interactions with it that could reveal where things went awry. "I wish I had a good story," Ariely told *Science*. "And I just don't."

Finding possible fraud in the work of such an influential scientist is jarring, Dimant says, especially for "the new generation of researchers who follow in his footsteps." Behavioral scientists Leif Nelson and Joseph Simmons, who exposed the apparent fraud via their blog Data Colada together with their colleague Uri Simonsohn, say a thorough, transparent investigation is needed. But given other universities' past reluctance to investigate their own researchers, they are skeptical that Duke will conduct one.

PHOTO: MENAHEM KAHANA/AFP/GETTY IMAGES

That may leave Ariely's supporters insisting he is innocent and detractors assuming he is guilty, Nelson says. "No one knows. And that's terrible."

The 2012 paper, published in the *Proceedings of the National Academy of Sciences* (*PNAS*), reported a field study for which an unnamed insurance company purportedly randomized 13,488 customers to sign an honesty declaration at either the top or bottom of a form asking for an update to their odometer reading. Those who signed at the top were more honest, according to the study: They reported driving 2428 miles (3907 kilometers) more on average than those who signed at the bottom, which would result in a higher insurance premium. The paper also contained data from two lab experiments showing similar results from upfront honesty declarations.

The Obama administration's Social and Behavioral Sciences Team recommended the intervention as a "nonfinancial incentive" to improve honesty, for instance on tax declarations, in its 2016 annual report. Lemonade, an insurance company, hired Ariely as its "chief behavioral officer." But several other studies found that an upfront honesty declaration did not lead people to be more truthful; one even concluded it led to more false claims.

After discovering the result didn't replicate in what he thought would be a "straightforward" extension study, one of the authors of the *PNAS* paper, Harvard Business School behavioral scientist Max Bazerman, asked the other authors to collaborate on a replication of one of their two lab experiments. This time, the team found no effects on honesty, it reported in 2020, again in *PNAS*.

While conducting the new lab study, Harvard Business School Ph.D. student Ariella Kristal found an odd detail in the original field study: Customers asked to sign at the top had significantly different baseline mileages—about 15,000 miles lower on average—than customers who signed at the bottom. The researchers reported this as a possible randomization failure in the 2020 paper, and also published the full data set.

Some time later, a group of anonymous researchers downloaded those data, according to last week's post on Data Colada. A simple look at the participants' mileage distribution revealed something very suspicious. Other data sets of people's driving distances show a bell curve, with some people driving a lot, a few very little, and most somewhere in the middle. In the 2012 study, there was an unusually equal spread: Roughly the same number of people drove every distance between 0 and 50,000 miles. "I was flabbergasted," says the researcher who made the discovery. (They spoke to *Science* on condition of anonymity because of fears for their career.)

Worrying that *PNAS* would not investigate the issue thoroughly, the whistleblower contacted the Data Colada bloggers instead, who conducted a follow-up review that convinced them the field study results were statistically impossible.

For example, a set of odometer readings provided by customers when they first signed up for insurance, apparently real, was duplicated to suggest the study had twice as many participants, with random numbers between one and 1000 added to the original mileages to disguise the deceit. In the spreadsheet, the original figures appeared in the font Calibri, but each had a close twin in another font, Cambria, with the same number of cars listed on the policy, and odometer readings within 1000 miles of the original. In 1 million simulated versions of the experiment, the same kind of similarity appeared not a single time, Simmons, Nelson, and Simonsohn found. "These data are not just excessively similar," they write. "They are impossibly similar."

Ariely calls the analysis "damning" and "clear beyond doubt." He says he has requested a retraction, as have his co-authors, separately. "We are aware of the situation and are in communication with the authors," *PNAS* Editorial Ethics Manager Yael Fitzpatrick said in a statement to *Science*.

Three of the authors say they were only involved in the two lab studies reported in the paper; a fourth, Boston University behavioral economist Nina Mazar, forwarded the Data Colada investigators a 16 February 2011 email from Ariely with an attached Excel file that contains the problems identified in the blog post. Its metadata suggest Ariely had created the file 3 days earlier.

Ariely tells *Science* he made a mistake in not checking the data he received from the insurance company, and that he no longer has the company's original file. He says Duke's integrity office told him the university's IT department does not have email records from that long ago. His contacts at the insurance company no longer work there, Ariely adds, but he is seeking someone at the company who could find archived emails or files that could clear his name. His publication of the full data set last year showed he was unaware of any problems with it, he says: "I'm not an idiot. This is a very easy fraud to catch."

Marc Ruef, an independent data forensics specialist, says Ariely could show as the "creator" of the Excel file even if the data did originate elsewhere, for instance because he created the spreadsheet and sent it to an insurance company to populate. But some behavioral scientists have asked on social media why a company would make up data about its clients' behavior in a way that supported one of Ariely's theories. (Ariely, citing Duke's legal advice, declined to name the company or comment about its involvement in possible fraud.)

The timeline is also hazy: Ariely mentioned the study in a 2008 lecture and in a 2009 *Harvard Business Review* piece, years before the metadata indicates the Excel file was created. Ariely says he does not remember when the study was conducted.

The odometer study has resurfaced other worries about Ariely's work. In July, an expression of concern was attached to a paper he published in 2004 in *Psychological Science*; in that case, statistical errors could not be resolved because Ariely was unable to produce the original data. In a 2010 NPR interview, Ariely referred to dental insurance data that the company involved later said did not exist, WBUR reported.

The Data Colada bloggers say they consider Ariely a friend. Finding his name as the creator of the field data file was "a very unpleasant moment," Simmons says. "This whole thing has been incredibly stressful." ■



"I wish I had a good story," says Duke University behavioral scientist Dan Ariely. "And I just don't."

PHOTO: LENGEMANN/WELT/ULLSTEIN BILD/GETTY IMAGES



Researchers plan to host new Mars colonization experiments in this pressurized, sealed greenhouse on the Biosphere 2 campus in Arizona.

PHOTO: TRENT TRESCH

ECOLOGY

Scaled down, martian model habitat rises again in desert

Leaders have big dreams for greenhouse at Biosphere 2

By **Michael Price**

There's a story, possibly apocryphal, about a cosmonaut who conducted a plant-growing experiment aboard a Russian orbiter in the 1970s or '80s. When the cosmonaut returned to Earth, he was evasive about the experiment's outcome. "I ate it," he finally declared. "I just had to eat something fresh."

Kai Staats, director of a new research facility known as the Space Analog for the Moon and Mars (SAM), likes to tell this tale to illustrate two things about space travel: It's hard to overestimate the importance of fresh vegetables when you're living in a tin can for months or years on end, and human beings are unpredictable.

SAM will attempt to address both concerns when it becomes operational later this year as the world's newest Mars analog, and strives to come closest to simulating a martian habitat. Such model space habitats—about a dozen around the world—aim to test technology and practice mission protocols, as well as mimic the harsh living conditions astronauts will face if they travel to the Red Planet decades from now. (NASA announced recently it plans to 3D print its own analog soon, and called for recruits to live there.) SAM is being refurbished from part of the most famous analog, the Biosphere 2 facility outside Tucson, Arizona, which next month marks its 30th anniversary.

SAM is a welcome addition, says ecologist Shannon Rupert, director of another Mars analog, the Utah-based Mars Desert Research Station. "SAM has two distinct blessings," she says. "First, it's already connected to a known destination, Biosphere 2, so it can have a huge public impact." Second, its airtight and pressurized facility "is the first of its kind." But the project—for now a single building—still needs funding to support its grander vision.

If it succeeds, SAM could redeem some of the promise of Biosphere 2, a sealed-ecosystem experiment that was part hippie commune, part serious scientific endeavor. Eight "biospherians" spent 2 years living in the contained, futuristic habitat, which featured acres of cropland, a rainforest, desert, and an indoor sea. But technological failures and crew conflicts overshadowed its mission; today it hosts earth science projects and is also a tourist attraction.

Staats, a documentary filmmaker and applied mathematician, worked with researchers from the University of Arizona to renovate a structure built in 1987 as a "test module" for the main Biosphere 2 building. About the size of a small house, the test module—which shares the futuristic, semigeodesic design of its big brother—will serve as SAM's airtight greenhouse. Eventually, Staats plans to connect it to three shipping containers with living quarters and lab space. All can be sealed off to maintain a positive internal air pressure to keep outside air from leaking in. In June, rehabbers pressurized the greenhouse for the first time in 30 years.

The original Biosphere 2 experiment infamously failed to achieve self-sustaining breathable air and needed external oxygen. Researchers have more modest goals for SAM: a system in which, over time, plants and humans produce and consume more and more of the oxygen and carbon dioxide (CO_2) needed to support one another.

The dynamic will change each time inhabitants harvest plants or sow seeds. Microbes add a further complication, as the original biospherians discovered when bacterial decomposition unexpectedly drove up CO_2 levels. Tracking the transition from imported air to air that has been recycled by plants inside SAM will help researchers model exactly which and how many plants and mushrooms to grow and harvest to maintain the balance, Staats says. "The goal is to learn how to start with all mechanical systems, as will be the case when we first land on the Moon or Mars, and then move toward plant-based systems."

Real-world data on the interplay of gases inside the greenhouse will be invaluable, says Martha Lenio, a renewable energy engineer at the World Wide Fund for Nature who served as mission commander for an 8-month jaunt at another analog. It's relatively easy to model the relationship between oxygen and CO_2 in an idealized setting, but "what do you do when things get out of whack?" she asks. "The human variable is not well understood. I think they're going to get a lot of useful information."

Staats envisions researchers from academia, industry, and space agencies paying to run experiments—say, on plant growth in sealed environments—and test space-bound equipment. People can live and work inside SAM for weeks or months at a time. The first clients arrive early next year; two were part of the original Biosphere crew.

Eventually, Staats plans to install a 2000-square-meter "Mars yard." But the facility is far from finished, and he doesn't have the needed funds. The team has worked on a shoestring budget, having received about $100,000 from private donors and a tech incubator. They'll need another $250,000 or so to fully equip the facility, Staats says. He hopes that money will come from investors, firms paying to use the facility, and grants.

Staats says this time around, the project's goals are achievable. "Biosphere 2 represents what we might do on Mars—[maybe]—100 or 150 years from now … [with] an ocean, a rainforest, all the things we're accustomed to at home. But SAM represents what we're going to do 20 to 25 years from now. We're scaling it back to what's realistic." ■

EARTH SCIENCE

Looking to the Rockies for clues to water woes

New effort to understand precipitation and runoff could improve Colorado River forecasts

By **Erik Stokstad**

In a historic first, the U.S. Bureau of Reclamation earlier this month declared a water shortage on the Colorado River, triggering emergency measures that will require farmers in Arizona to cut their use of irrigation water by 20% next year. The immediate cause of the declaration is record low water levels in Lake Mead, the largest reservoir fed by the river. But scientists say the crisis has been years in the making—and could soon get worse. For reasons they don't completely understand, but that are related to the West's changing climate, snow that falls in the Rocky Mountains—the source of about 80% of the Colorado—has been providing the river with less and less water. "This is an existential water crisis for the Southwest," says Jonathan Overpeck, a climate scientist at the University of Michigan, Ann Arbor.

Next week, researchers will begin an innovative campaign to better understand the fundamental processes—from the behavior of tiny particles that become snowflakes to weather patterns that influence how snow vanishes into thin air—that determine how mountain precipitation becomes surface water for 40 million people. "What gets us going in the morning is the large number of people that really rely on this resource," says atmospheric scientist Daniel Feldman of the Department of Energy's (DOE's) Lawrence Berkeley National Laboratory (LBNL), who leads the effort.

For the more than $8 million project, called the Surface Atmosphere Integrated Field Laboratory (SAIL), researchers are deploying dozens of instruments that will measure wind, rain, snow, solar radiation, and atmospheric particles in a high-elevation Colorado watershed. Hydrologists have already been studying the streams and bedrock there for years. But the additional equipment will collect data intended to sharpen models that produce a variety of critical forecasts, including short-term predictions of seasonal stream flows and long-term scenarios of how climate change might alter regional water supplies. SAIL is "going to make advances in mountain precipitation and snow studies that would just be impossible without this level of instrumentation," says Jessica Lundquist, a mountain hydrologist at the University of Washington, Seattle. "It's really exciting."

The southwestern United States has been in the grips of a drought for 2 decades (*Science*, 2 July, p. 17). A drop in precipitation has been exacerbated by a decline in what's called runoff efficiency, or the proportion of precipitation that reaches waterways. In the 1930s and '40s, about 17% of basinwide precipitation ended up in the Colorado River, scientists say. Now, it's roughly 14%, a decline that has contributed significantly to a recent 20% loss of river flow. An overall warming of the West seems to be driving the change, and possible mechanisms include increased water uptake by plants and a loss of snow cover, which means the ground and atmosphere become warmer because less sunlight is reflected back into space. Even though the trend in runoff efficiency is clear, computer models have had a hard time predicting exactly how much water will reach rivers, in part because the physical mechanisms involved are murky.



Gothic Mountain looms over a major new water research project in Colorado.

PHOTO: U.S. DEPARTMENT OF ENERGY ATMOSPHERIC RADIATION MEASUREMENT USER FACILITY

SAIL's technicians have been working all summer in Colorado's 300-square-kilometer East River watershed, installing instruments that are part of DOE's mobile Atmospheric Radiation Measurement program. To get a better view of the rain and snow, they have installed one radar system at the top of a ski resort. It has a resolution of 100 meters, five to 10 times greater than a typical weather radar, says V. Chandrasekar of Colorado State University. It will allow researchers to identify and quantify precipitation, including various types of snow particles, that are up to 50 kilometers away.

Other instruments should enable SAIL to observe hard-to-track interactions, such as those involving the tiny aerosol particles that help create snowflakes. Researchers also hope to probe the mysteries of a process called sublimation, in which snow turns directly into water vapor. In certain conditions, sublimation claims about 20% of the peak snowpack before it melts. Sublimation has rarely been studied in combination with so many other atmospheric processes, says Jeff Lukas, a water and climate researcher and consultant based in Colorado. SAIL will also measure where, when, and how much rain and snow fall, helping hydrologists better understand how water moves through mountains and into streams, says co–principal investigator Rosemary Carroll, a hydrologist at the Desert Research Institute.

In parallel with SAIL, the National Oceanographic and Atmospheric Administration is mounting an effort to improve its weather models and mountain region forecasts. And the U.S. Geological Survey will install new stream gauges to Colorado headwater streams as part of its Next Generation Water Observing System.

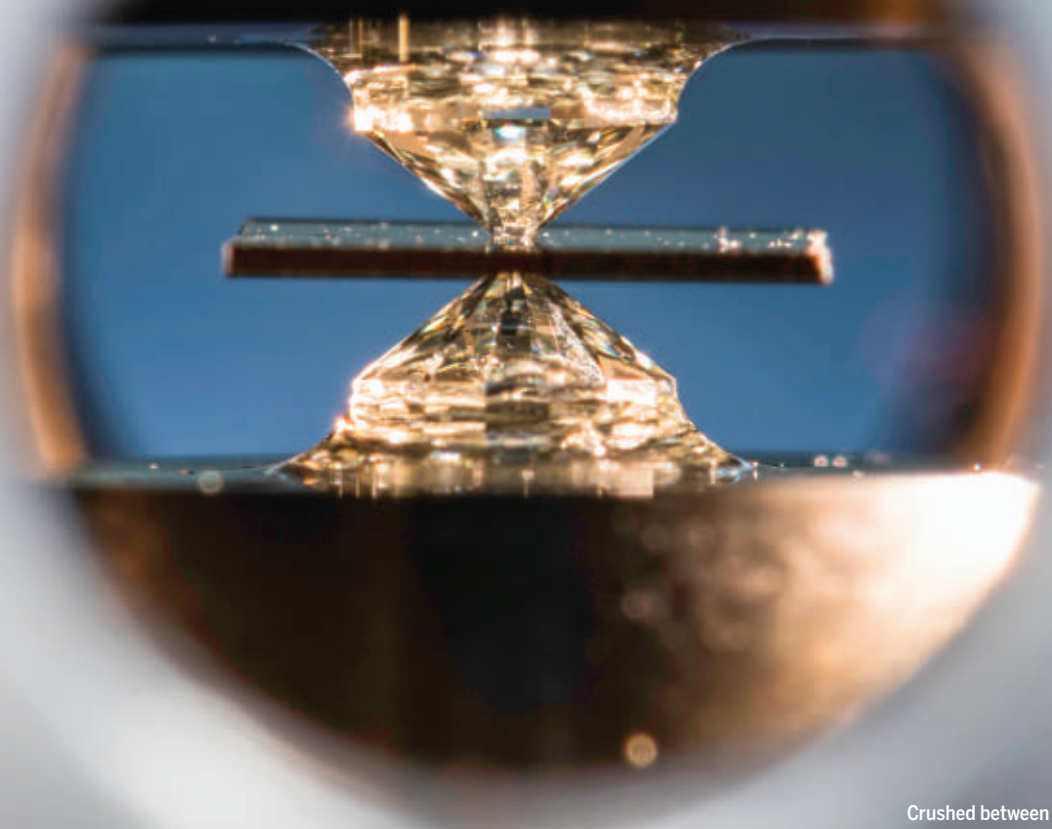
The campaign's biggest synergy, however, is with a major hydrology and biogeochemistry investigation that DOE has funded in the East River watershed since 2014, examining how it stores and releases water, carbon, nutrients, and pollution. Collaborations there now involve hundreds of scientists, notes LBNL's Susan Hubbard, who leads the project. And that diversity thrills many SAIL participants. "It's like a science rodeo," says ecohydrologist Alejandro Flores of Boise State University.

One key question is how easy it will be to transfer any insights gained from SAIL to other mountain regions around the world where rivers are facing similar water supply issues. "If we don't understand all the details, we're going into this water management challenge with one arm tied behind our back," Overpeck says. "Possibly more like both arms." ■

FEELING THE PRESSURE

Can room-temperature superconductors transform from a tantalizing glimmer to practical materials?

By **Robert F. Service**, *in Rochester, New York*



Crushed between two tiny gems in this diamond anvil cell, a compound of carbon, sulfur, and hydrogen showed signs of superconductivity at room temperature.

PHOTO: J. ADAM FENSTER/UNIVERSITY OF ROCHESTER

Every year, Ranga Dias pulverizes about $100,000 in diamonds, crushing them to a gray powder he tosses in the trash. It's worth it. The superhard gems were sacrificed to achieve a goal researchers have chased for generations.

Before their demise, pairs of the 2-millimeter-diameter gems serve as jaws in a miniature vise. In fall 2020, Dias, a physicist at the University of Rochester (U of R), and colleagues used the setup to squash a few specks of carbon and sulfur along with a whiff of hydrogen gas to a pressure near that found at Earth's center. The force rearranged the elements into carbon sulfur hydride (CSH), reported to be the first substance that can conduct electricity with no electrical losses at room temperature—a chilly room, to be sure. That experiment ended the same way nearly all such efforts do: The diamond jaws exploded, and the world's only room-temperature superconductor vanished in a puff.

To date, no one else has matched the feat. But other high-pressure physics groups are crushing their own diamonds in the attempt. And those researchers have found several other hydrogen-rich materials that superconduct at relatively warm temperatures. Collectively known as hydrides, the cousins contain hydrogen along with other elements, such as lanthanum or yttrium, and still need to be chilled a little more (or sometimes a lot more) than Dias's CSH.

Together, the hydrides have revolutionized superconductivity, showing it can happen at temperatures only dreamed of in the past, when superconductors revealed their powers only when cooled below –100°C. But so far, the materials can only be made in fractions of grams and only superconduct when squeezed to outlandish pressures. That limitation makes them wholly impractical for real-world applications. So a new goal has emerged: room-temperature superconductors that retain their magic when the pressure is off. They could enable advances as varied as a class of supercomputers that don't waste vast amounts of energy as heat or transmission lines that carry massive electrical loads without the usual losses, saving billions of dollars and megatons of carbon emissions.

"This will be a challenge," says Steven Louie, a superconductivity theorist at the University of California (UC), Berkeley, who doesn't work on hydrides. "But these are very exciting developments. For a long time, people thought high-temperature superconductors couldn't reach room temperature." The hydrides, he says, "show this is not true."

But not everyone buys the results. "It is easy to be fooled. These are difficult experiments done on extremely small samples under intense pressure," says Jorge Hirsch, a theoretical physicist at UC San Diego who is a leading critic of claims for hydride superconductivity. "I agree something is going on. But I think it's not going to turn out to be interesting new physics, but an experimental artifact."

"Any time this kind of paradigm shift happens, people will push back," Dias responds. He agrees with Hirsch that high-pressure experiments are difficult to pull off and can produce results that are hard to interpret. But Dias and fellow hydride researchers are confident in their observations. And with theoretical predictions and preliminary experimental results both suggesting that dramatically lowering the need for pressure might be possible, the scientists press forward.

> **"People thought high-temperature superconductors couldn't reach room temperature. ... [Hydrides] show this is not true."**
>
> **Steven Louie**, University of California, Berkeley

DESPITE THE DIAMONDS, Dias's lab is no jewel box. The facility sits on the first floor of a blocky four-story, 1960s-era building. The lab's heart is a fist-size, copper-colored cylinder called a diamond anvil cell. Once loaded with starting materials and screwed shut, the cell slots into a small blue metal box bolted to a table. The box houses a cryogenic cooler that uses liquid helium to precisely control the temperature. Mirrors on the table steer laser light through windows into the diamond anvil cell.

To get a superconductor, the researchers use a pressurized stream of inert gas to drive the diamonds together, generating a force amplified manyfold at their tips. A green laser aimed through the diamonds triggers chemical reactions that combine carbon, sulfur, and hydrogen atoms into a crystalline solid. The same laser then probes the material with Raman spectroscopy, which reveals which chemical elements are bonded to one another. And tiny wires positioned between the diamonds track the material's electrical resistance, to detect superconductivity's hallmark sharp plunge to zero. The experiments run nonstop—sometimes for weeks—until the inevitable cracking sound, when another pair of diamonds turns to dust.

Dias, a native of Sri Lanka, moved to the United States to pursue his Ph.D. at Washington State University, studying how explosions can shock materials into new forms. From there, he moved to Harvard University to do a postdoc with Isaac Silvera, a physicist working to use intense pressures to turn hydrogen gas into a metal, which they suspected might be superconducting. Other researchers had previously claimed to have seen signs of metallic hydrogen, but those results remained controversial. In 2017 in *Science*, Dias and Silvera said they had succeeded by trapping hydrogen in a diamond anvil cell and dialing the pressure up to 495 gigapascals (GPa), more than 4.8 million times atmospheric pressure, transforming the gas into a silvery solid.

That result generated plenty of praise—and pushback. "The word garbage cannot really describe it," another expert said at the time. Critics cautioned that, among other things, the Harvard pair only published a single result, without replication. But Dias and Silvera have stuck by their results. And Silvera says he hopes to publish additional positive results soon.

Even if metallic hydrogen is real and confirmed to be superconducting, a substance that only survives at millions of atmospheres of pressure won't make much of a real-world impact. That's why proving superconductivity in hydrides—which show promise for superconductivity at lower pressures—is a much more important challenge. "It would have a huge impact," Dias says.

DISCOVERED IN 1911 by Dutch physicist Heike Kamerlingh Onnes, the original superconductors were chunks of elemental metals, such as mercury and niobium, cooled to a few degrees above absolute zero. At such frigid temperatures, electrical resistance vanishes because electrons no longer bump into atoms as they travel and give up a bit of their energy to heat.

In 1933, German physicist Walther Meissner noted an added feature of superconductors: They expel magnetic fields. The phenomenon remained a mystery until 1957, when physicists John Bardeen, Leon Cooper, and Robert Schrieffer explained those first superconductors. Their "BCS theory" suggested an electron whizzing through a superconducting metal deforms the material's atomic lattice, drawing positive atomic nuclei toward it ever so slightly. That redistribution of charges around an electron pulls another electron behind it, akin to how the weight of one person on a bed draws a second person toward them. That effect explains the resistance-free flow. The exclusion of magnetic fields, known as the Meissner effect, is a byproduct of the free-flowing electrons: When an external magnetic field meets a

superconductor, the field induces an electric current to pirouette within the superconductor, which generates its own magnetic field that cancels out the external field.

The next major advance came in 1986, when J. Georg Bednorz and K. Alexander Müller, physicists in IBM's Zurich Research Laboratory, discovered that copper oxide ceramics, known as cuprates, could superconduct at a relatively balmy 30 K. That "critical temperature," or T_c, was soon pushed up in other cuprates to above 77 K, a temperature achievable with liquid nitrogen. That finding opened the door to broader applications because liquid nitrogen is far cheaper than the liquid hydrogen needed to chill the first superconductors. By 2005, researchers had pushed the T_c as high as 166 K (or −107°C), under high pressure, with a mercury-based cuprate. But then the march up the thermometer stopped.

Another potential route to high-temperature superconductors beckoned: hydrogen. In 1968, Neil Ashcroft, a theoretical physicist at Cornell University, suggested putting hydrogen under intense pressure would turn the gas into a solid lattice able to superconduct. His ideas languished because vises weren't yet producing the necessary pressures; not until design refinements came along could groups including Silvera's try to make solid metallic hydrogen. And then a new line of research opened up.

In 2004, Ashcroft suggested that adding other elements to hydrogen might add a "chemical precompression," stabilizing the hydrogen lattice at lower pressures. The race was on to make superconducting hydrides. In 2015, researchers including Mikhail Eremets, a physicist at the Max Planck Institute for Chemistry, reported in *Nature* that a mix of sulfur and hydrogen superconducted at 203 K when pressurized to 155 GPa. Over the next 3 years, Eremets and others boosted the T_c as high as 250 K in hydrides containing the heavy metal lanthanum. Then came Dias's CSH compound, reported late last year in *Nature*, which superconducts at 287 K—or 14°C, the temperature of a wine cellar—under 267 GPa of pressure, followed by an yttrium hydride that superconducts at nearly as warm a temperature, announced by multiple groups this year. "It's all moving very fast," says Lilia Boeri, a theoretical physicist at the Sapienza University of Rome.

TOO FAST, SOME RESEARCHERS argue. Hirsch and others caution that the high-pressure results don't actually show a key feature of superconductors: the exclusion of magnetic fields. No one has yet figured out how to measure the effect inside a diamond vise. "They have experiments that involve clever techniques to show the magnetic field is reduced inside," says Marvin Cohen, a theoretical physicist at UC Berkeley. But that's not good enough for many scientists in the field. "I'm old fashioned," says Paul Chu, a physicist at the University of Houston and a veteran of battles over superconductivity in the cuprates. "I always like to see the Meissner effect. I haven't seen it."

Another problem is the resistance data, Hirsch says. When an external magnetic field is applied to a superconductor, the T_c typically drops, with a greater drop for stronger magnetic fields. In a graph of temperature versus electrical resistance, the curve normally showing the drop in electrical resistance at T_c flattens out as the external field increases. And higher temperature superconductors typically show a flatter slope. But in the hydrides, the highest temperature superconductors of all, "they don't see this," Hirsch says. "Either this is a nonstandard superconductor, or it is not a superconductor."

Eremets, whose team has discovered multiple hydride superconductors, says Hirsch is wrong, at least about the hydrides Eremets has worked on. His group's 2019 *Nature Communications* paper on a hydrogen sulfide superconductor contains a plot showing the T_c drops as an external magnetic field is increased, meaning the resistance flattens as expected for a superconductor. "Simply, he missed it," Eremets says.

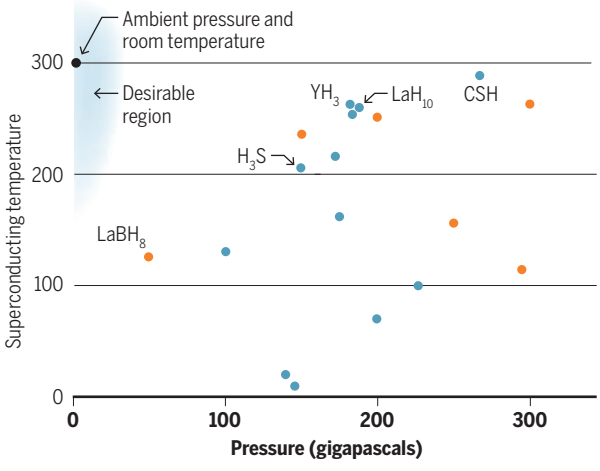
Dias adds that in some hydrides, the resistance may not behave as expected because the tiny samples in the diamond cell may be highly purified. He says the resistance curve flattening in other superconductors as an external magnetic field increases is actually an artifact of impurities, which can affect a material's properties. Highly pure superconductors, such as the classic superconductor magnesium diboride, show a sharper drop. "Our results are in line with MgB_2," Dias says.

But Hirsch has another concern about Dias's CSH material: Its magnetic susceptibility, a measure of how much a material becomes magnetized in an applied magnetic field, does not behave like other superconductors'. To test magnetic susceptibility, researchers apply a magnetic field to a potential superconductor as it is cooled. In standard superconductors, magnetic susceptibility drops as the material is cooled below T_c and stays low as the temperature continues to drop. Dias, however, reported that CSH's magnetic susceptibility drops as the material cools below T_c but then rises again as cooling continues. "A superconductor doesn't do that," says Hirsch, who concludes that either the data are wrong or the material isn't superconducting. He has asked the authors, the editors at *Nature*, and the National Science Foundation, which funded the work, to supply raw data, but so far has not been given access. "This has me very frustrated," Hirsch says.

Dias says his team is filing patent applications, so his lawyers have asked him to withhold the data for now. He adds that the apparent rise in magnetic susceptibility below T_c is an artifact. That increase isn't unusual in

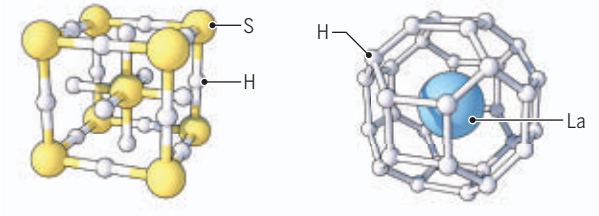
New materials, long-sought goal

Using high pressures, experimentalists have created many superconducting hydride compounds (●), including one—carbon sulfur hydride (CSH)—that appears to work at close to room temperature. Theorists have predicted other hydride compounds (●). Now, the race is on to find versions stable at ambient pressure and room temperature.



Hydride structures

The atomic structure hasn't been determined for all hydride superconductors. But so far they fall into two general classes, with examples shown below.



Sulfur hydride (H_3S)
In this group of hydrides, hydrogen atoms (white) bind directly to other light atoms, such as sulfur (yellow).

Lanthanum hydride (LaH_{10})
In this class, hydrogen atoms (white) form a cage surrounding electron-rich atoms, such as lanthanum (blue).

CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) NATHAN DASENBROCK-GAMMON/UNIVERSITY OF ROCHESTER

high-pressure experiments, Dias says, because they produce a background signal that can overwhelm the experimental signal.

Dias and others concede one of their critics' points: So far, the Meissner effect is impossible to show inside a vise. However, the researchers say a technique that measures a property known as AC susceptibility accomplishes much the same thing. The technique involves using tiny magnetic coils next to the sample to create an oscillating magnetic field and then watching for any induced voltage change in the material. As superconductors are cooled below T_c, the voltage typically shows a drop. And the hydrides do exactly that. "All the evidence we are seeing is in line with the BCS model," Dias says.

Eva Zurek, a theoretical chemist at the University at Buffalo, and Louie are confident hydrides will pass muster as superconductors, noting that multiple groups have reported evidence of superconductivity in sulfur hydride (H_3S), lanthanum hydride, and yttrium hydride. "To me, that's convincing," Zurek says.

However, even other hydride experimentalists are hesitant to sign off on Dias's room-temperature superconductor, which remains a one-off result. "We've been trying to replicate this for the past 6 months, but so far we haven't seen it," Eremets says. Katsuya Shimizu, a physicist at Osaka University who has made H_3S, is also cautious. "I don't believe it before I repeat it," Shimizu says. Dias urges patience. "It will take time to resolve all this," he says. "It took years for us to get there. You can't expect others to get there in a few weeks."

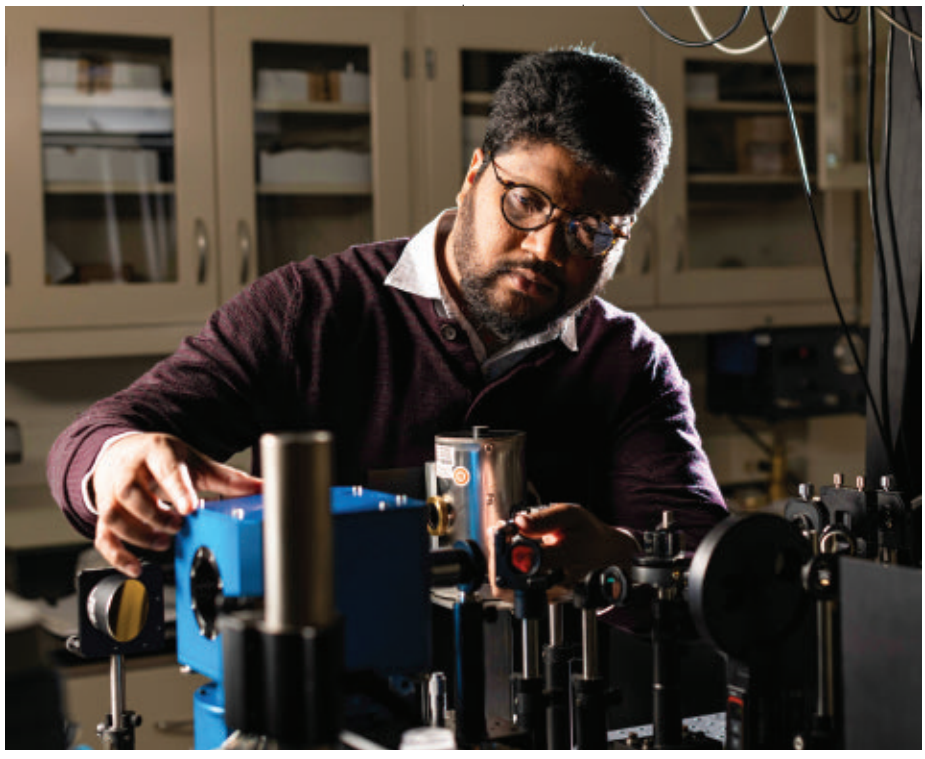


PHOTO: J. ADAM FENSTER/UNIVERSITY OF ROCHESTER

Physicist Ranga Dias pursues his quest for new superconductors in a diamond anvil cell, housed in the blue box. His goal, however, is a room-temperature superconductor that would survive outside the cell.

IF THE CSH RESULTS do hold up, that raises an important question: "Is it possible to get to ambient [pressure]?" Dias asks. "That's what we're all pushing for now."

The first order of business will be to find other hydrides that superconduct at room temperature. In normal superconductors, T_c depends on two main factors. First is the abundance of electrons in the material that aren't stuck to individual atoms but are free to conduct. Typically, the more conducting electrons, the better. Second is how fast atoms in the crystalline lattice vibrate—the lattice vibrations essentially bind pairs of electrons, and faster is better. Hydrogen, the lightest atom, vibrates fastest, but a lattice of hydrogen atoms isn't very rigid and easily falls to bits. So scientists have to apply external pressure to keep it from flying apart.

To raise T_c and lower pressure, researchers need chemical recipes that either add a bunch of electrons to the hydrogen lattice or lock hydrogen into a stiffer lattice. Researchers have reported success with both strategies, leading to two classes of hydrides with very different 3D structures. The first class includes repeating cagelike structures made from hydrogen atoms, with each cage enclosing an electron-rich metal atom, such as lanthanum or yttrium. The second class adds light elements designed to bind directly with hydrogen to create a continuous network of interlocking atoms (see graphic, p. 956).

Dias and colleagues, including Suxing Hu, a theoretical physicist at U of R, believe CSH forms such an interlocking grid. In a 13 May preprint, researchers including Russell Hemley, a chemist at the University of Illinois, Chicago, shone x-rays at a CSH sample made by Dias's group in a diamond cell. Hemley's group saw evidence of a regular crystallization pattern of sulfur atoms—at least up to 178 GPa, when the diamonds broke. Hu and colleagues then used Hemley's data to model a structure for CSH. The team found that CSH most likely adopts the same cubelike structure as H_3S, which superconducts at a temperature 80 K lower than CSH does. Because carbon tends to form stronger bonds to its neighbors than sulfur or hydrogen do, carbon may be what's holding the lattice together at room temperature.

To reduce pressure further, Hu, Dias, and others suggest adding more carbon, or perhaps boron, another strong bonder. In fact, Boeri and colleagues published a paper in *Physical Review B* on 15 July predicting that lanthanum borohydride ($LaBH_8$) could superconduct at 126 K under just 50 GPa of pressure. Other theorists have predicted that hydrides such as calcium hydride or actinium hydride should superconduct at close to room temperature—and at a pressure considerably less than that needed for CSH. Still, Boeri says, "I'm not sure we can get to ambient pressure."

Dias is more upbeat. At the March meeting of the American Physical Society, he said he had preliminary evidence for a room-temperature hydride superconductor that is stable down to 20 GPa, less than one-tenth the pressure CSH required. But because he and his team are patenting that discovery, too, and collecting more data, he's unwilling to say what the material is. If it holds up, he predicts, "it will be a significant step forward. I think we'll get there soon."

He's banking on it. He recently formed a company called Unearthly Materials to manufacture and commercialize hydride superconductors, if any can be found to work near ambient pressures. If researchers can ease the pressure a bit more, to below 10 GPa, they can ditch the diamond anvil cells. The door would then be open to making larger samples of hydrides, allowing research groups to enter the fray even without access to cells, and enabling straightforward probes for the Meissner effect, the superconductivity signal that skeptics are still waiting to see.

What began with crushed diamonds could turn into a gold rush. ■



PHOTO: BOB WICK/BUREAU OF LAND MANAGEMENT

ENVIRONMENTAL POLICY

Allow "nonuse rights" to conserve natural resources

"Use-it-or-lose-it" requirements should be reconsidered

By **Bryan Leonard[1,2], Shawn Regan[2], Christopher Costello[2,3,4], Suzi Kerr[5], Dominic P. Parker[2,6,7], Andrew J. Plantinga[3], James Salzman[3,8], V. Kerry Smith[1,4,9], Temple Stoellinger[10,11]**

Market approaches to environmental conservation, by which mechanisms such as property rights, prices, and contracts are used to advance environmental goals, have gained traction globally in recent decades (*1*). But in many cases, antiquated rules limit their role in conserving public natural resources. "Use-it-or-lose-it" requirements, together with narrow definitions of eligible "uses," can preclude environmental groups from participating in markets for natural resources. These restrictions can bias resource management in favor of extractive users, even when conservation interests are willing to pay more to protect resources from development. We argue that acquisition of public natural resource rights for the purpose of withholding them from development should be allowed. Policies should be reformed to include conservation as a legally valid form of "use." Allowing such "nonuse rights" to public natural resources would enable markets to advance environmental goals, leading to more stable and less contentious outcomes.

Use-it-or-lose-it rules are legally mandated for many publicly managed resources, meaning that rights to the resource are granted on the condition that the resource be exploited. For example, at least 16 of 37

Nonuse rights have been thwarted on public lands in areas such as southern Utah, shown here.

Organisation for Economic Co-operation and Development (OECD) countries have use-it-or-lose-it provisions for water allocation (*2*). In the United States, these requirements were established in the 19th and early 20th centuries when the nation's primary management goal was to promote productive use of natural resources, defined as extraction. Such rules are outdated today because the demands for alternative, conservation-oriented "uses"—recreation, protection of ecosystem services, and scenic views—often exceed extractive demands for the same resources.

A recent example illustrates the problem. In 2016, environmental activist Terry Tempest Williams purchased drilling rights to 450 ha of federal land in Utah for $2500 at a Bureau of Land Management (BLM) lease auction. To qualify as a bidder, she created a company—Tempest Exploration Co. LLC—and began paying rental fees on the lease. But when Tempest Williams revealed that she intended to keep the oil in the ground, the BLM canceled the leases, arguing that she violated the "diligent development requirement" of the 1920 Mineral Leasing Act, which requires lessees to "exercise reasonable diligence in developing and producing" their leases.

Such requirements may hinder several current US and international conservation goals. Under an executive order issued soon after taking office (EO 14008), the Biden administration announced plans to conserve 30% of US lands and waters by 2030 (referred to as "30 by 30"). To meet that goal, EO 14008 also called for a "comprehensive review and reconsideration" of the federal oil and gas leasing program. Recent scientific proposals to address climate change, protect biodiversity, and safeguard ecosystem services (*3*) have prompted many nations to propose similar goals.

Although we explore this theme here in the context of US law and policy, our discussion is relevant internationally, where public management is common for surface resources, subsurface deposits, fisheries, water, and pollution permits. For example, the supply-side approach of "keep it in the ground" for fossil fuels requires countries to allow nonuse of commercially valuable public resources (*4*). Moreover, the growing international trend toward recognizing environmental flows as a valid use of water suggests that countries often draw on the experiences of other nations when crafting their own definitions of acceptable "uses" of natural resources (*2*). Nonuse rights are also central to ongoing debates over extending legal rights to nature, as in the case of New Zealand's Whanganui River, which was granted legal personhood in 2017. They also complement efforts to restore resource governance authority to Indigenous communities, which often place greater emphasis on nonextractive resource uses.

THE STATUS QUO

With the exception of national parks, designated wilderness, and other protected areas—which together make up only about one-third of US federal lands—the rules governing public natural resources are biased toward extractive uses, requiring leaseholders to extract, graze, divert, harvest, or otherwise develop resources (see the box). These rules made sense more than a century ago to discourage waste and prevent speculation, but they create new challenges today. Efficient resource management requires a more modern interpretation of natural resource "use." Changing values, better scientific information about goods and services provided by intact ecosystems, and income growth have fueled new demands for conservation on public lands that are not formally classified as protected. Meanwhile, "unused" landscapes are becoming scarcer. Natural resource policy has been slow to respond to these changes. This may be due in part to the entrenched power of constituencies who benefit from extractive uses and therefore resist policy changes that threaten their livelihoods.

Because environmental nongovernmental organizations (ENGOs) generally cannot acquire public resource rights directly, environmental values are largely expressed through regulations that define permissible uses or through litigation under various environmental statutes. Rather than directly acquiring natural resources and conserving them, ENGOs must instead expend considerable resources attempting to influence political, legal, or administrative processes, often with little or only short-lived success.

Moreover, political and legal approaches often pit extraction-dependent communities against environmentalists and can result in contentious outcomes that are vulnerable to shifts in partisan control. In its final year, the Trump administration reversed several management policies of prior administrations such as allowing logging in previously protected areas of Alaska's Tongass National Forest (*5*). Though ENGOs opposed these policies—and likely would have paid more than traditional users for leases to conserve these areas—they lacked the ability to acquire lease rights. Although the Biden administration may undo some of these policy changes, a future administration could reverse course yet again.

Expanding public natural resource rights to include conservation will require amending or reinterpreting US laws that govern resource management, some more than a century old. Depending on the resource, this could be done by an act of Congress or state legislatures, by revisions to administrative rules by the executive branch, or sometimes through judicial rulings. Because administrative rules can be modified relatively easily by future administrations, legislative reforms provide the most reliable way to create nonuse rights that are reasonably secure and free from political interference.

A CASE FOR NONUSE RIGHTS

If done with care, letting conservationists acquire resource rights has the potential to improve the status quo for several reasons. First, markets can reveal information about the economic values for what are currently considered nonuses of public resources. Measures for the values of recreation and conservation are sensitive to assumptions about the extent of the market and the degree of substitutability between different resources and sites (*6*). Allowing market exchanges of resource rights for both use and nonuse purposes could reveal how the value of additional conservation differs across locations and circumstances, enabling ENGOs to identify and conserve areas where there is substantial unmet conservation demand.

Second, there is evidence that ENGOs would participate in markets to conserve public natural resources if allowed to do so. There are active markets for conservation easements and payments for ecosystem services on private land—where nonuse rights to natural resources are recognized (*7*). In instances where ENGOs have been allowed to acquire nonuse rights to publicly managed resources—under either region-specific legislation that authorizes nonuse acquisitions, administrative procedures to enable nonuse buyouts at an agency's discretion, or state requirements to accept high bidders for leases on state-owned lands—several groups have demonstrated a willing-

[1]School of Sustainability, Arizona State University, Tempe, AZ, USA. [2]Property and Environment Research Center, Bozeman, MT, USA. [3]Bren School of Environmental Science and Management, University of California, Santa Barbara, CA, USA. [4]National Bureau of Economic Research, Cambridge, MA, USA. [5]Environmental Defense Fund, Washington, DC, USA. [6]Department of Agricultural and Applied Economics, University of Wisconsin–Madison, Madison, WI, USA. [7]Hoover Institution at Stanford University, Stanford, CA, USA. [8]UCLA School of Law, University of California, Los Angeles, CA, USA. [9]Department of Economics, Arizona State University, Tempe, AZ, USA. [10]Haub School of Environment and Natural Resources, University of Wyoming, Laramie, WY, USA. [11]College of Law, University of Wyoming, Laramie, WY, USA. Email: bryan.leonard@asu.edu

ness to acquire these rights (*8*). ENGOS have purchased federal energy leases, negotiated voluntary grazing permit retirements with ranchers, contracted to leave water instream for fish, and outbid logging companies for timber leases (see table S1). The authority to acquire nonuse rights currently applies to only a few regions and specific circumstances but could be expanded to other areas of public natural resource management.

Third, nonuse rights have the potential to deliver more secure and lasting conservation outcomes than traditional management policies, as long as such rights are reasonably secure and well defined. In the 1990s, the Grand Canyon Trust, an Arizona-based ENGO, negotiated agreements with ranchers to relinquish their federal grazing permits in Utah's Grand Staircase-Escalante National Monument. Even though the ENGO successfully struck a deal with permittees, there was no certainty that the BLM would not later reissue the permits to other ranchers. Because of federal grazing requirements, the ENGO could not hold the permits themselves without grazing livestock (*8*). Indeed, in 2020, the US Department of the Interior announced plans to reopen these retired allotments to grazing. By contrast, if the rules allowed the group to acquire the permits for conservation purposes, such rights would be secure through clear contractual obligations and would have well-defined time horizons.

Fourth, rights-based approaches allow conservation priorities to adapt to evolving market and environmental conditions. When relative values change or new information emerges, market participants can respond quickly by buying or selling existing rights (*9*). These adjustments reflect the natural arbitrage we would expect as extraction and conservation demands respond to environmental change. For example, climate change may alter the variability, timing, and spatial distribution of wildlife migrations, wildfires, and streamflows, among others (*10*).

Fifth, market approaches can be designed to compensate current extractive users who stand to lose if policy priorities and economic demands change. ENGOs, for example, could be allowed to negotiate buyouts of existing leaseholders' rights to public natural resources (*8*). More than 10.5 million ha of US public lands are currently leased for oil and gas production—half of which are not yet developed but could be in the near future—and more than 87 million ha are leased for livestock grazing. Letting ENGOs purchase existing leases for conservation purposes would provide a way for current leaseholders to be compensated for relinquishing their permits. Recent state-level reforms that have enabled ENGOs to acquire water rights for environmental purposes are one potential model. Purchases of instream flow rights by Trout Unlimited, Freshwater Trust, and other ENGOs have secured critical fish habitat while compensating traditional water users for the associated reductions. This strategy is likely to receive more widespread support than curtailing extractive users' rights through regulation (*11*).

Legal barriers to conservation leasing of US public natural resources

Grazing
- The Taylor Grazing Act of 1934 gives preference for grazing permits to those who reside within or near a designated grazing district and are engaged in the livestock business.
- If permittees do not graze at or near the authorized level, permits can be revoked and transferred to another permittee.
- Permittees must own or lease qualifying private property that can serve as the base for a livestock operation.

Energy and Minerals
- The General Mining Act of 1872 provides mineral prospectors who discover hard rock minerals an exclusive "right to mine." Claims are maintained perpetually through an annual $100 maintenance fee or labor and improvements at the claim site worth $100 each year.
- The Mineral Leasing Act of 1920 requires oil and gas lessees to "exercise reasonable diligence in developing and producing" energy resources. If resources are not developed within the 10-year primary term, the lease can be terminated and made available for other developers.

Timber
- The National Forest Management Act of 1976 requires that the terms of timber sale contracts "shall be designated to promote orderly harvesting."
- Timber leases on national forests are must-cut contracts. Failure to cut timber within a designated length of time (not to exceed 10 years) voids the contract.

Water
- The prior appropriation system imposes a use-it-or-lose-it requirement on water rights holders in western states. Water must be put to certain defined "beneficial uses," which historically excluded nonuse conservation. Some states now consider conservation, or instream flow, to be a beneficial use, although others do not.

Fisheries
- In rights-based fisheries, quota holders are generally required to own vessels and/or harvest or lease their quota.

Wildlife
- Rights to wild game are only established by harvesting an individual animal rather than by owning a license to hunt. Nonharvested game is under the control of state wildlife agencies and is often available to be harvested by others.

RESEARCH PRIORITIES FOR AVOIDING PITFALLS

Market-based policies can have unintended consequences if not approached with care. For example, stakeholders from local communities may oppose allowing ENGOs to buy out extractive users because of potential economic losses from curtailed development. In the case of public lands, nonuse rights could create challenges for communities and industries that are reliant on activities related to drilling, grazing, or logging operations by reducing regional demand for labor and inputs associated with these activities (*12*). At the same time, there is some evidence that nonextractive uses of public lands (e.g., recreation and complementary preservation) can spur economic growth for rural communities (*13*). In any case, large-scale buyouts for conservation could bring substantial economic and cultural shifts for some communities, and nonuse rights should be designed and implemented in ways that address the concerns of existing stakeholders.

Proliferation of nonuse rights may create revenue challenges and management obstacles for federal and state governments. Although allowing environmental bidders could result in higher revenues from lease auctions, governments may be required to forego royalties accrued from the extraction of resources such as oil, gas, and timber. Bids from extractive users and ENGOs may not be directly comparable because an extractive user may offer a lower up-front bid but provide governments more revenue over time from royalty payments. Nonuse rights may also create challenges for governments whose management strategies are intertwined with existing uses. In some cases,

extraction may be necessary for long-term resource management. Livestock grazing and timber harvesting, for example, can play a beneficial role in resource management and in reducing wildfire risk by limiting the build-up of fuels (*14*).

Carefully constructed rules for nonuse rights could help overcome some of these hurdles. For example, agencies could price in foregone royalties or require higher annual rental payments from nonuse lessees. Agencies could also specify management outcomes that must be achieved regardless of who holds the resource rights. For resources such as timber and rangelands, agencies could mandate wildfire mitigation measures or require lessees to achieve certain ecological outcomes, similar to how the US Forest Service has used stewardship contracting and the BLM has used outcome-based grazing authorizations to meet desired resource conditions. For resources such as wildlife and fisheries, this may be more difficult if managers use harvests (e.g., hunting or fishing) to achieve specific population targets.

Another concern is that nonuse rights, if acquired on a large scale in one area, could create political pressure for managers to make additional resources available for extraction elsewhere, which in turn might offset some of the conservation gains from nonuse acquisitions. This could be prevented by rules that constrain the ability of managers to offer more resources for use in response to acquisitions by ENGOs. For example, states that allow water rights for instream flows often set minimum-flow requirements for specific streams and limit the authority of agencies to grant new water rights that would reduce flows below those standards. Moreover, to the extent that nonuse acquisitions shift extractive activities away from areas with high conservation value and toward areas with low conservation value, such substitution would be consistent with a more efficient spatial allocation of conservation and extraction, even if there were no net decrease in extraction.

Yet another concern is that developers could exploit nonuse rights to create monopoly power by buying up large amounts of extraction rights and withholding them from production. States have addressed this issue in the context of nonuse rights for water through various "antispeculation" measures that prevent instream flow rights from becoming a source of monopoly power. Some fisheries in the United States, Canada, and Iceland have "consolidation caps" that limit the amount of tradable fishing quota that can be controlled by a single entity. Similar rules could be adopted for other resources. Concerns over market power also vary by resource. Locally priced resources like water are more vulnerable to monopoly control than globally traded commodities like oil, timber, and beef.

Spatial externalities could affect the demand for nonuse rights if, for example, access roads, oil and gas infrastructure, or wandering cattle spill onto adjacent areas leased by ENGOs. As a practical matter, a combination of administrative rules and local nuisance and property law would govern such conflicts. More broadly, the scope for these spillovers would likely cause ENGOs to acquire large contiguous areas, as has been the case when exceptions to existing rules allowed nonuse rights.

Experience from existing environmental markets suggests that many of these concerns can be mitigated through careful design and implementation, guided by specific research. Recent social science insights on opposition to rights-based management by resource-dependent communities should be extended and combined with emerging work on gradual or "just" transitions to ensure that communities' concerns are fully addressed (*15*). Insights from mechanism design and auction theory could be applied to address questions of contract structure, revenue generation, and ENGOs' incentives to participate that may vary across different resources (e.g., nonrenewable oil and gas versus renewable timber stands). Finally, our evolving understanding of the ecological impacts of different human uses of land can inform the assessment of how expanded nonuse rights can realize management targets.

Beyond concerns over design and implementation, some may simply object to the idea of having to pay to conserve public natural resources. Yet agencies have longstanding responsibilities to allow multiple uses—including for both extraction and conservation purposes—and to generate fair returns to taxpayers or share revenues with nearby communities. Some state land agencies are even required to maximize resource revenues to benefit schools and other public institutions. Not allowing ENGOs to bid creates an implicit subsidy to extractive users, who face less competition and hence lower prices. Nonuse rights would expand opportunities for conservation in a way that acknowledges multiple-use management responsibilities, respects the rights of existing users, and reflects the opportunity costs of foregone extraction.

Market-based approaches to conservation are not a panacea. Private provision of environmental public goods will always be subject to some degree of free riding, and the magnitude of the conservation benefits that could be achieved with nonuse rights remains an open question. Free riding also affects the informational benefits of market mechanisms. Still, if ENGOs successfully acquire nonuse rights, then the purchase price provides a lower bound on conservation values. Moreover, the benefits of nonuse rights will no doubt be greater in some contexts than in others depending on the environmental services and stakeholders involved. For example, ENGOs may have more difficulty raising money to keep oil and gas in the ground as a strategy to reduce global carbon emissions than to preserve locally prized amenities. At the same time, the rise of crowdsourcing has reduced the transaction costs of coordinating buyers in ways that could enable national, if not global, participation to fund resource conservation.

Ultimately, the benefits and costs of extending market approaches to conservation must be compared against the shortcomings of the status-quo reliance on political and administrative mechanisms. Well-crafted rights-based approaches can help avoid the controversy that has hamstrung previous attempts to advance conservation of public natural resources. As the Biden administration reconsiders its federal oil and gas leasing program and promulgates rules to advance large-scale conservation of US lands and waters, it should embrace markets for nonuse rights to address growing demands for the conservation of public natural resources. ■

REFERENCES AND NOTES

1. T. L. Anderson, G. D. Libecap, *Environmental Markets: A Property Rights Approach* (Cambridge Univ. Press, 2014).
2. OECD, Water Resources Allocation: Sharing Risks and Opportunities (OECD Studies on Water, OECD Publishing, Paris, 2015).
3. E. Dinerstein *et al*., *Sci. Adv*. **5**, eaaw2869 (2019).
4. G. B. Asheim *et al*., *Science* **365**, 325 (2019).
5. Harvard Law School, Environmental and Energy Law Program, "Regulatory tracker"; https://eelp.law.harvard.edu/regulatory-rollback-tracker/.
6. R. Mendelsohn, S. Olmstead, *Annu. Rev. Environ. Resour*. **34**, 325 (2009).
7. J. Salzman, G. Bennett, N. Carroll, A. Goldstein, M. Jenkins, *Nat. Sustain*. **1**, 136 (2018).
8. B. Leonard, S. Regan, *Nat. Resour. J*. **59**, 135 (2019).
9. S. E. Anderson *et al*., *Clim. Change Econ*. **10**, 1950003 (2019).
10. G. T. Pecl *et al*., *Science* **355**, eaai9214 (2017).
11. J. C. Neuman, *Neb. Law Rev*. **432**, 484 (2004).
12. H. Eichman *et al*., *J. Agric. Resour. Econ*. **35**, 316 (2010).
13. M. Walls, P. Lee, M. Ashenfarb, *Sci. Adv*. **6**, eaay8523 (2020).
14. M. Cochrane *et al*., *Int. J. Wildland Fire* **21**, 357 (2012).
15. P. Newell, D. Mulvaney, *Geogr. J*. **179**, 132 (2013).

ACKNOWLEDGMENTS

The Property and Environment Research Center provided funding to host working group meetings to facilitate the writing of this manuscript.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/958/suppl/DC1

10.1126/science.abi4573

PERSPECTIVES

SUPERCONDUCTIVITY

Driving multiphase superconductivity

Breaking local symmetry opens up a clear transition between superconducting states

By **Alexandre Pourret** and **Georg Knebel**

Symmetry and symmetry breaking are keys to much of the interesting phenomena in condensed matter physics. Conventional superconductivity, for example, requires both time reversal and inversion symmetry, and the removal of one of these (such as time reversal through a magnetic field) leads to the suppression of the superconducting order. Recently, there is a realization that even if the global symmetry of the system is present, a local symmetry breaking can still induce a variety of fascinating behaviors. On page 1012 of this issue, S. Khim *et al.* (*1*) report on field-induced transition within the superconducting state of $CeRh_2As_2$ driven by local inversion-symmetry breaking.

In a three-dimensional superconductor, inversion and time-reversal symmetry are the only symmetries that ensure that a fermionic state with momentum **k** is energy degenerate with another state at **–k**. This degeneracy is necessary to support a weak-coupling superconducting instability because of the formation of Cooper pairs with total momentum **q** = 0. Consequently, these two symmetries are at the basis of the usual classification of even-parity (spin-singlet) and odd-parity (spin-triplet) Cooper pairs that result from the Pauli exclusion principle, which requires a completely antisymmetric wave function for indistinguishable fermions. In non-centrosymmetric heavy-fermion superconductors (*2*)—such as $CePt_3Si$ (*3*), $CeIrSi_3$ (*4*), $CeCoGe_3$ (*5*), and $CeRhSi_3$ (*6*)—the lack of inversion symmetry introduces a mixing of even and odd parity pairing states. These mixed states generally reveal no Pauli paramagnetic limit. This leads to what is probably the most spectacular experimental result, the observation of extremely high upper critical fields for a field applied in the direction lacking mirror symmetry—that is, perpendicular to the basal plane (see the figure).

The heavy fermion material Khim *et al.* studied, $CeRh_2As_2$, is a variant of non-centrosymmetric superconductors. The crystal structure is globally centrosymmetric (with well-defined inversion centers) but is locally non-centrosymmetric, with an inversion symmetry linking two non-centrosymmetric Ce-square lattices. This configuration can be compared with the relation between ferromagnetic (FM) and antiferromagnetic (AF) structures. The FM state globally breaks time-reversal symmetry; the AF state does not but has two sublattices of subunits violating time-reversal symmetry. The consequence of such a structure in $CeRh_2As_2$ is that even-parity (spin-singlet) and odd-parity (spin-triplet) Cooper pairs are not mixed, so that a phase transition between even- and odd-parity condensates should occur inside the superconducting condensate.

S. Khim *et al* observed a clear phase transition between two superconducting phases for field parallel to the **c** axis. The observation of multiple superconducting phases is restricted to only a few materials. The theoretical approach to model the superconducting state involves two different channels of the pairing interaction based on a spin-singlet pairing between electrons on the same Ce sublattice. The first pairing channel implies an even-parity state in which the Cooper pair wave function on these two sublattices has the same sign, and the second one is the odd-parity state in which the Cooper pair wave function has opposite sign on the different sublattices. The even-parity channel is expected to have the higher superconducting transition temperature and is also more strongly suppressed by a Zeeman field, which allows the odd-parity solution

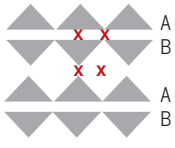
A local departure from global symmetry

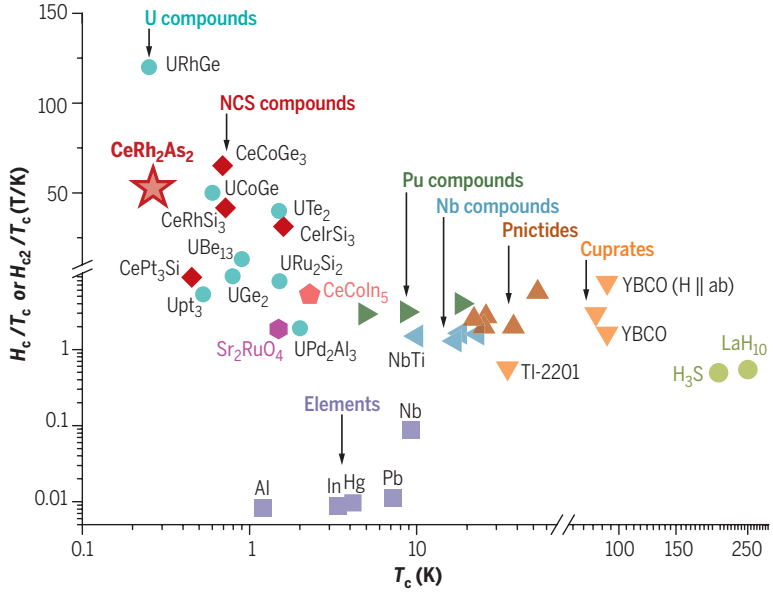
$CeRh_2As_2$ is among the non-centrosymmetric superconductors (NCSs) that feature a high critical magnetic field (H_c or H_{c2}) divided by the superconducting temperature (T_c). The high value at low T_c comes out of the distinct crystal structure.

Layered globally non-centrosymmetric
The layers are identical and do not have an inversion center.



Layered staggered non-centrosymmetric
$CeRh_2As_2$ has global inversion centers (red crosses) but a A/B substack structure that is locally non-centrosymmetric.





GRAPHIC: N. CARY/*SCIENCE*

Université Grenoble Alpes, Commissariat à l'Energie Atomique et aux Energies Alternatives, Grenoble Insititute of Technology, Interdisciplinary Research Institute of Grenoble, Laboratoire Photonique Electronique et Ingénierie Quantiques, 38000, Grenoble, France.
Email: alexandre.pourret@cea.fr

to appear at higher fields. The competition between these channels is controlled by details of the electronic structure. In particular, when the Rashba spin-orbit coupling (SOC) because of the locally broken inversion symmetry is larger than the intersublattice hopping, this gives rise to near-degenerate even- and odd-parity pairing states and a substantial enhancement of the superconducting upper critical field (*7*).

The discovery by the authors of a transition between two distinct superconducting phases under the application of a magnetic field in $CeRh_2As_2$ coincides with a renewal of interest in odd-parity superconducting pairing states. Similar transition has been observed under pressure inside the superconducting phase (*8*) of the recently discovered heavy fermion superconductor UTe_2 (*9*, *10*). In contrast to $CeRh_2As_2$, the superconducting phases in UTe_2 are believed to be all spin triplet, and rather than Rashba spin-orbit coupling, spin fluctuations are thought to be the driving force of the relevant physics. Beside the pairing mechanism being different, the presence of spin triplet pairing in both systems is at the origin of similar unconventional features such as the very high upper critical field, anisotropy of the upper critical field, and multiple superconducting phases.

"...odd-parity superconductors may play a central role for quantum materials..."

Today, odd-parity superconductors may play a central role for quantum materials science because they can host nontrivial topological phenomena. In quantum engineering science, a strong effort has been made to induce odd-parity superconductivity in ferromagnetic materials through proximity effect in nanoscale devices. The discovery of new quantum materials—such as $CeRh_2As_2$, in which such possibly topological superconducting states appear naturally in bulk—creates an opportunity to study new emergent phenomena in condensed matter physics. ■

REFERENCES AND NOTES

1. S. Khim *et al.*, *Science* **373**, 1012 (2021).
2. M. Sigrist *et al.*, *J. Phys. Soc. Jpn.* **83**, 061014 (2014).
3. E. Bauer *et al.*, *Phys. Rev. Lett.* **92**, 027003 (2004).
4. T. Akazawa *et al.*, *J. Phys. Soc. Jpn.* **73**, 3129 (2004).
5. A. Thamizhavel *et al.*, *J. Phys. Soc. Jpn.* **74**, 1858 (2005).
6. N. Kimura *et al.*, *Phys. Rev. Lett.* **95**, 247004 (2005).
7. D. C. Cavanagh, T. Shishidou, M. Weinert, P. M. R. Brydon, D. F. Agterberg, arXiv:2106.02698 [cond-mat.supr-con] (2021).
8. D. Braithwaite *et al.*, *Commun. Phys.* **2**, 147 (2019).
9. S. Ran *et al.*, *Science* **365**, 684 (2019).
10. D. Aoki *et al.*, *J. Phys. Soc. Jpn.* **88**, 043702 (2019).

10.1126/science.abj8193

THERMAL TRANSPORT

Pushing low thermal conductivity to the limit

Weak bonding between the distorted layer of $Bi_4O_4SeCl_2$ helps limit its heat transport

By **Shi En Kim**[1] *and* **David G. Cahill**[2]

A wide variety of materials with low thermal conductivity find daily use, such as jackets for cold weather and plastic handles of hot metal cooking pots. Even the best thermal insulators still have a finite thermal conductivity because the vibrational motion of atoms is never fully localized and energy is transported through coherent collective vibrations (phonons). Any thermal excitation that is not fully localized can carry heat, which adds to the challenge of realizing materials with ultralow thermal conductivity for real-world applications. On page 1017 of this issue, Gibson *et al.* (*1*) have an answer to how low thermal conductivity can go. They synthesized layered inorganic bulk crystals and measured a thermal conductivity that is an order of magnitude lower than that of typical oxide glasses and only four times the value of air.

The discovery of materials with extreme thermal conductivities is important for many applications, especially in energy systems. Just as dissipation of electrical energy is absent in a perfect electrical insulator or a superconductor, the generation of entropy and the attendant loss of efficiency in thermal processes would be zero for a hypothetical perfect thermal insulator or conductor. Microelectronic systems would also benefit from improved materials. For example, thermally insulating inorganic crystals would be useful for confining heat in thermal devices such as phase-change memory (*2*).

Using solid-state chemistry techniques, Gibson *et al.* synthesized highly anisotropic, ordered-disordered structures in the layered van der Waal crystals $Bi_2O_2Cl_2$ and Bi_2O_2Se, as well as the naturally occurring superlattice $Bi_4O_4SeCl_2$. Instead of controlling the orderliness of atoms in the material (a common route of materials engineering), they manipulated the bond strength and connectivity between interlayer atoms. In crystals, phonon modes channel heat with higher efficiency compared with atoms moving independently. Gibson *et al.* designed bonding motifs to manipulate the travel velocities of phonons. By tuning interfaces of their material and the choice of the unit cell, they could select for the types of phonons their crystals would let through to produce an exceedingly low thermal conductivity.

The inorganic crystals Gibson *et al.* synthesized join the legion of other materials with extremely low thermal conductivities (see the figure) (*3*). The most notable of these materials are disordered layered transition metal dichalcogenide (TMD) crystals, in which polycrystalline layers are misaligned with respect to the adjacent layers (*4*, *5*). These exotic materials undershoot the lower thermal conductivity regime at which scientists once thought thermal conductivity to be "poor" by dipping below the amorphous limit (*6*) designated to materials with similar composition and density but with random atomic placements and bonding.

All of these exotic materials that beat lower bounds have a nonintuitive mix of atomic order and disorder in crystals. Previous studies have shown that the deliberate amorphization of the randomly rotated structure of misaligned TMD stacks only raised the thermal conductivity. Thus, some other mechanism not captured by traditional thermal transport models operates in these inorganic crystals.

Gibson *et al.* used a suite of material design parameters that address the failure of traditional theory to universally describe low–thermal conductivity materials. Their strategy was to fabricate ultralow–thermal conductivity crystals with a blend of anharmonicity, anisotropy, and partial atomic disorder. Their material defies any particular category of low–thermal conductivity materials that are better understood, such as fully disordered harmonic glasses (*7*) and strongly disordered crystals (*8*). Yet, even the theoretical descriptions of these established material groups suffer from gaps. Although perfect crystals with weak anharmonicity are well understood (*9*), strong anharmonicity that drives the phonon mean

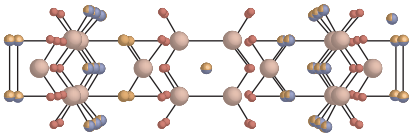
[1]Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, USA. [2]Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61822, USA. Email: d-cahill@illinois.edu

Strategies for ultralow thermal conductivity

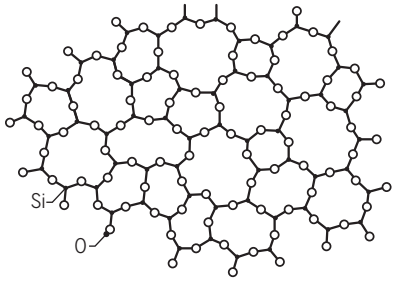
Different types of atomic arrangements in materials can interfere with transport of coherent lattice vibrations (phonons) and lead to low thermal conductivity. In the material developed by Gibson *et al.*, all three effects are incorporated.

Bi Cl Tl O Al
Si V K Se H

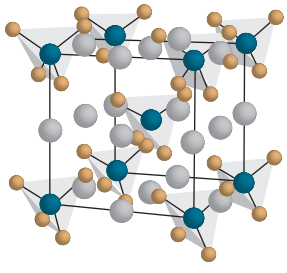
Interface engineering
Gibson *et al.* manipulated the bond strength and connectivity between interlayer atoms in $Bi_4O_4SeCl_2$.



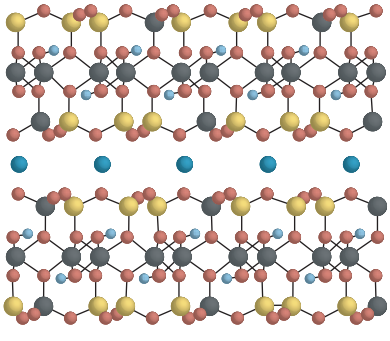
Disorder in atomic arrangements
Phonons undergo more scattering in amorphous materials such as ordinary glass (silicon dioxide) than in ordered crystals.



Anharmonicity
Phonon modes do not propagate well if the vibrations are anharmonic (excited modes are not exact multiples of the ground state mode). The structure in Tl_3VSe_4 is unstable with respect to transforming to a different bonding geometry.



Anisotropic structures
Strong bonding within planes and weak bonding across planes combine to suppress heat conduction, as shown for mica.



free path down to the wavelength of the phonon itself makes the assumptions of the theory invalid. A recent example is Tl_3VSe_4 (*10*). The difficulty in describing phonon transport in such materials has practical implications. For example, it impairs assessing the potential efficiency of materials as thermoelectrics for interconversion of heat and electrical power.

Recent research by Simoncelli *et al.* (*11*) elucidated low–thermal conductivity transport phenomena where traditional models fail. They developed a phonon thermal conductivity model to unify the well-developed descriptions for harmonic glasses and for weakly anharmonic crystals and then applied this approach to understanding the low thermal conductivity of the isotropic perovskite $CsPbBr_3$. Their approach could guide experimentalists trying to find materials with even lower thermal conductivities. A critical test of the validity of this approach is challenging, however. Thermal conductivity integrates heat carried by a broad spectrum of thermally excited vibrations. Thus, thermal transport measurements are not sensitive to these individual spectral features for any theory to be rigorously tested.

A fully validated and comprehensive theory that bridges the entire thermal conductivity spectrum and is universally applicable to most, if not all, materials, is still lacking. For example, the theoretical understanding of thermal conductivity transport in disordered soft materials, such as amorphous and liquid crystalline polymers, is incomplete. Like many previous studies, Gibson's *et al.*'s explanation for the crystals' ultralow thermal conductivity may work well in this particular instance but may not be generalizable. For low–thermal conductivity materials in general, theory and experiment need to be reconciled, and the limited theoretical understanding of thermal transport remains at a bottleneck even as the list of exotic thermal conductors grows. ■

REFERENCES AND NOTES

1. Q. D. Gibson *et al.*, *Science* **373**, 1017 (2021).
2. H. Kwon *et al.*, *Nano Lett.* **21**, 5984 (2021).
3. M. Beekman, D. G. Cahill, *Cryst. Res. Technol.* **52**, 1700114 (2017).
4. C. Chiritescu *et al.*, *Science* **315**, 351 (2007).
5. D. Li, A. Schleife, D. G. Cahill, G. Mitchson, D. C. Johnson, *Phys. Rev. Mater.* **3**, 043607 (2019).
6. G. A. Slack, *Thermal Conductivity of Non-metallic Crystals, in Solid State Physics*, F. Seitz, D. Turnbull, Eds. (Academic Press, 1979), vol. 34, p. 57.
7. J. L. Feldman, M. D. Kluge, P. B. Allen, F. Wooten, *Phys. Rev. B Condens. Matter* **48**, 12589 (1993).
8. D. G. Cahill, S. K. Watson, R. O. Pohl, *Phys. Rev. B Condens. Matter* **46**, 6131 (1992).
9. L. Lindsay, A. Katre, A. Cepellotti, N. Mingo, *J. Appl. Phys.* **126**, 050902 (2019).
10. Z. Zeng *et al.*, *Phys. Rev. B* **103**, 224307 (2021).
11. M. Simoncelli, N. Marzari, F. Mauri, *Nat. Phys.* **15**, 809 (2019).

10.1126/science.abk1176

MOLECULAR BIOLOGY

Piercing the fog of the RNA structure-ome

Machine learning is poised to transform RNA structure and function discovery

By **Kevin M. Weeks**

RNA is distinct among large biomolecules in that it has both informational coding ability, carried in its sequence, and the ability to form complex three-dimensional structures that can have catalytic and regulatory roles. The information-carrying component is widely appreciated. The pattern of base pairing—the first level of RNA structure—can be experimentally assessed and modeled with impressive accuracy (*1*, *2*). By contrast, our understanding of the extent and roles of complex three-dimensional RNA structures remains rudimentary. RNA viral genomes are rich in motifs with complex three-dimensional structures with regulatory functions (*3*), and evidence increasingly supports the hypothesis that functional RNA structures are ubiquitous in organisms ranging from bacteria to humans. However, developing and testing hypotheses about the roles of RNA structure have been hindered by the inability to identify and model these structures. On page 1047 of this issue, Townshend *et al.* (*4*) report a machine-learning strategy for identifying native-like RNA folds.

Nearly all RNAs that form well-understood complex structures fall into a small number of classes: the ribosomal RNAs, the large and small ribozymes that catalyze RNA cleavage, bacterial riboswitches, and regulatory elements encoded by RNA viruses. Thus, there are limited examples for guiding identification and modeling of RNAs with complex three-dimensional structures. There are only four major RNA nucleotides, and the interactions that govern base pairing and simple helix formation are well understood. Once formed, RNA helices (secondary structure) often assemble as fairly rigid elements that interact hierarchically to form more complicated structures (tertiary structure) (see the figure). Despite these simplifying features,

Department of Chemistry, University of North Carolina, Chapel Hill, NC, USA. Email: weeks@unc.edu

GRAPHIC: N. DESAI/*SCIENCE*

the modeling of complex RNA structures has proven to be difficult.

The RNA-Puzzles community exercise (*5*, *6*) has been instrumental in illuminating the challenges involved: Groups try to predict an RNA structure from its sequence before learning the solved structure. Several rounds of RNA-Puzzles have revealed important themes. No single method consistently yields the best models, although certain approaches have better records than others, and most approaches are getting better. The best agreement tends to result when experimental or homology-based information is incorporated into the computational modeling. However, the median accuracy for small RNAs, with complex tertiary folds but without a close known homolog, has stayed stubbornly stuck in a range of ~15- to 20-Å root mean square deviation [(RMSD) a measure of the similarity between known and modeled structures]. This agreement is much poorer than that now achieved for protein structures by machine learning (*7*), where native-like folds (~2-Å RMSD or less) are achieved. Modeled RNA structures thus often recapitulate the overall fold of a target RNA but do not consistently reveal details of the tertiary structure. Current methods are not likely to be useful for applications such as understanding the biological mechanism of a structure or for designing ligands (or drugs) that modulate RNA function.

The Atomic Rotationally Equivalent Scorer (ARES) approach of Townshend *et al.* is a deep neural network, a form of machine learning, and did not initially include preconceived notions of RNA structure. Indeed, the ARES framework is not specific to RNA and can be applied to other problems in molecular structure. Instead, ARES was given a small set of motifs with known RNA structure plus a large number of alternative (incorrect) variations of these same structures. ARES parameters were adjusted so that the program learned the functional and geometric arrangements of each atom and how these elements are positioned relative to each other. Layers in the neural network compute features from finer to coarser scales to recognize base pairs, helices, and more-complex structures. For example, ARES learned patterns of base pairing, the optimal geometry for RNA helices, and a subset of noncanonical tertiary motifs without being provided explicit information about these features of RNA structure.

Although ARES was trained on very simple RNA systems, the resulting ARES scoring function was able to predict structures of more complex RNAs, on average, to roughly a 12-Å RMSD. This degree of accuracy represents an overall improvement of ~4 Å over prior scoring methods. ARES is still short of the level consistent with atomic resolution or sufficient to guide identification of key functional sites or drug discovery efforts, but Townshend *et al.* have achieved notable progress in a field that has proven recalcitrant to transformative advances.

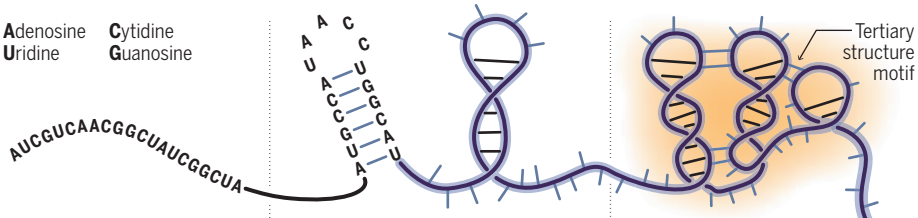
There are three fundamental challenges for modeling complex RNA three-dimensional structures: generating reasonable structures that may represent a biological state, accurately scoring or identifying models that best represent the correct native state, and using these hopefully accurate models to discover new functional motifs and to develop hypotheses regarding the mechanisms by which RNAs with complex three-dimensional structures regulate biological processes. The ARES machine-learning approach addressed the second of these three challenges: Candidate structures still need to be generated for evaluation by ARES. With further development, deep learning strategies hold promise for creating new scoring functions that can guide structure generation in ways that might yield near-native structures. Another important goal is to use a machine-learning strategy to identify regions in large RNAs most likely to fold into three-dimensional structures.

Current computational-only algorithms are not able to predict the pattern of base pairing in large RNAs accurately, even though base pairs are simpler to predict than tertiary structure. However, secondary structures for large RNAs are routinely modeled to high accuracies by incorporating experimental information. New, efficiently executed experiments are now being developed that measure features of RNA tertiary structures. Another frontier, analogous to recent advances in secondary structure modeling, would thus be to incorporate experimental information into machine-learning strategies for modeling RNA tertiary structure.

Large-scale investigation of RNA structure to date, primarily focused on RNA secondary structure, has revealed several core principles. One is that the existence of regions within large RNAs with complex, higher-order structure is unremarkable. When these base pairing and tertiary structures affect biological functions, they create "an RNA structure code" with pervasive effects on gene regulatory circuits. Additionally, every RNA likely has a distinct structural personality, which implies that there are numerous ways by which RNA structure tunes the underlying function of an RNA. At the level of secondary structure, such tuning RNA structures tend to function like switches and attenuators that modulate binding by RNA and protein ligands (*8*–*11*). Finally, characterization of well-determined RNA secondary structures often leads to identification of centers of new biology. As it becomes possible to measure, (deeply) learn, and predict the details of the tertiary RNA structure-ome, diverse new discoveries in biological mechanisms await. ■

RNA structure

RNA molecules have multiple levels of structure and ability to encode information. The sequence of RNA is readily determined. RNA secondary structure can now be elucidated with high levels of accuracy using approaches that meld computational energy minimization with experimental per-nucleotide chemical probing information. Townshend *et al*. developed a deep neural network that can identify models that best represent the native tertiary state, taking a step toward modeling three-dimensional RNA structure.



Primary structure
The RNA sequence can encode a protein or serve as a binding site for single-stranded binding proteins or interacting RNAs (such as microRNAs).

Secondary structure
RNA secondary structures (helices) create functional motifs on their own and function as switches that alter the accessibility of constituent sequences.

Tertiary structure
Multiple secondary structures form higher-order (tertiary) motifs that can function as catalysts, ligand-binding domains, switches, and environmental sensors.

GRAPHIC: C. BICKEL/*SCIENCE*

REFERENCES AND NOTES

1. E. J. Strobel *et al.*, *Nat. Rev. Genet.* **19**, 615 (2018).
2. K. M. Weeks, *Acc. Chem. Res.* **54**, 2502 (2021).
3. Z. A. Jaafar, J. S. Kieft, *Nat. Rev. Microbiol.* **17**, 110 (2019).
4. R. J. L. Townshend *et al.*, *Science* **373**, 1047 (2021).
5. J. A. Cruz *et al.*, *RNA* **18**, 610 (2012).
6. Z. Miao *et al.*, *RNA* **26**, 982 (2020).
7. E. Pennisi, *Science* **373**, 262 (2021).
8. D. Long *et al.*, *Nat. Struct. Mol. Biol.* **14**, 287 (2007).
9. M. Kertesz *et al.*, *Nat. Genet.* **39**, 1278 (2007).
10. D. Dominguez *et al.*, *Mol. Cell* **70**, 854 (2018).
11. A. M. Mustoe *et al.*, *Biochemistry* **57**, 3537 (2018).

ACKNOWLEDGMENTS

The author's laboratory is supported by the US National Institutes of Health and National Science Foundation. The author is an advisor to and holds equity in Ribometrix.

10.1126/science.abk1971

MICROBIOLOGY

How microbiota improve immunotherapy

Ligands derived from the gut microbiota enhance cancer immunotherapy

By **Eduard Ansaldo**[1] and **Yasmine Belkaid**[1,2]

Animals have coevolved with complex communities of microorganisms living on barrier tissues, referred to as the microbiota or commensals. The microbiota controls immune function not only locally within barrier tissues but also systemically, modulating functions such as hematopoiesis, immune system development, and responses to vaccines (*1*). The role of the microbiota in enhancing responses to cancer immunotherapy has represented a major focus of research (*2*–*7*), although the mechanisms have remained largely unclear. On page 1040 of this issue, Griffin *et al.* (*8*) show that members of the *Enterococcus* genus promote immunotherapy responses in mice through immunostimulatory muropeptides, which are structural units of bacterial cell walls.

One of the most convincing pieces of evidence for the adjuvant role of the microbiota in immunotherapy came from recent clinical studies demonstrating that transplantation of fecal microbiota from melanoma patients responding to programmed cell death protein 1 (PD-1) immune checkpoint therapy, but not from nonresponders, improved the efficacy of these PD-1 inhibitors in melanoma patients who were previously refractory to therapy (*9*, *10*). Because of the extraordinary diversity of the microbiota, identification of defined mechanisms represents an enormous challenge, and the causal microbes or pathways identified so far have differed (*3*–*7*).

Griffin *et al.* focus on the genus *Enterococcus*, one of the taxa described to correlate with responses in patients treated with immune checkpoint therapy (*5*, *6*). The authors identified multiple species within this genus that are able to promote responses to PD-1 ligand 1 (PD-L1) immunotherapy in mouse tumor models. Notably, immunotherapy-active enterococci included the common human commensal *Enterococcus faecium*, whereas other species such as *E. faecalis* provided no protection. The authors were guided by previous studies that characterized the distinct composition of *E. faecium* peptidoglycan, the major structural component in bacterial cell walls. Comparison of the structure of peptidoglycans across immunotherapy-active and -inactive enterococci revealed defined peptidoglycan patterns that correlated with increased response to immunotherapy.

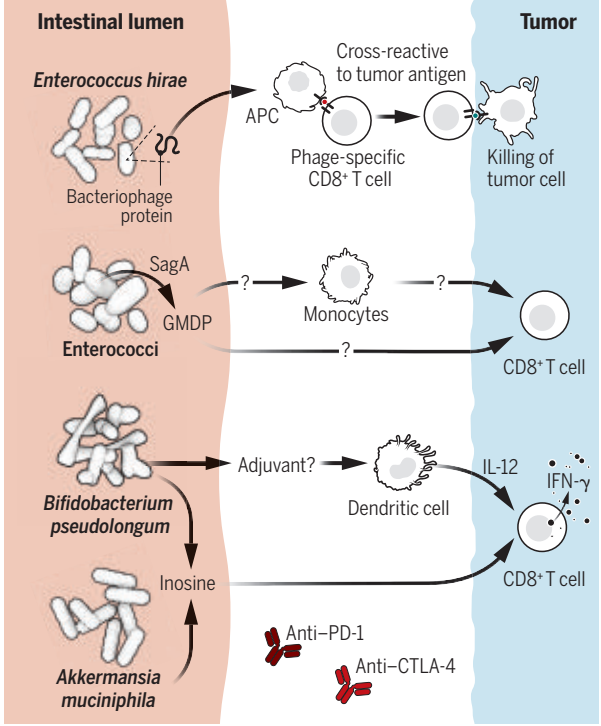
Using a comparative genomic analysis, Griffin *et al.* identified a cluster of peptidoglycan hydrolases conserved across immunotherapy-active species, including the peptidoglycan hydrolase secreted antigen A (SagA) that was sufficient to confer responses when ectopically expressed in *E. faecalis*. This was dependent on the innate immune sensor nucleotide-binding oligomerization domain–containing protein 2 (NOD2), which recognizes peptidoglycan-derived muropeptides. Of interest, NOD ligands have been associated with immune modulating effects of the microbiota, including hematopoiesis, response to vaccination, and susceptibility to Crohn's disease (*11*–*13*).

Griffin *et al.* take a critical step toward understanding the mechanism by which specific commensal species can promote responses to immunotherapy, but it is still unclear on which cell types NOD2 ligands are acting and how these lead to increased antitumor immunity. Although NOD1 is broadly expressed, NOD2 expression is restricted to immune cells and certain populations of nonhematopoietic cells such as intestinal epithelial cells (*13*). One possibility is that NOD2 ligands also act on myeloid cells in the tumor microenvironment and its draining lymph node to promote anticancer adaptive immunity. Of note, a recent study revealed that NOD2 ligand-based therapeutics modulated myelopoiesis in the bone marrow, leading to epigenetic rewiring of myeloid cells that were then able to overcome the immunosuppressive tumor microenvironment (*14*).

Griffin *et al.* expand on a small number of studies that were able to provide a molecular link between a microbe and its antitumor effects in the context of immune checkpoint blockade (*7*, *15*) (see the figure). A previous study identified commensal bacteria–derived inosine as a key determinant of *Bifidobacterium pseudolongum* and *Akkermansia muciniphila* to enhance the antitumor efficacy of immune checkpoint blockade through the promotion of type 1 immunity (*7*), a class of T cell–mediated immune responses associated with protective responses to pathogens. Similarly, Griffin *et al.* show that *Enterococcus*-derived peptidoglycan promotes the accumulation of cytotoxic CD8$^+$ T cells and associated innate responses in the tumor microenvironment. A similar link between the microbiota and increased type 1 immunity has been observed in preclinical models

Microbiota enhance immunotherapy

Microbiota can influence the response to cancer immunotherapy through multiple mechanisms. These include phage antigens that are cross-reactive with tumor cells, cell wall–derived muropeptides such as glutaminylmuramyl-dipeptide (GMDP), and commensal-derived inosine, all of which converge on increased type 1 immunity against the tumor in the context of immune checkpoint blockade therapy.



APC, antigen-presenting cell; CTLA-4, cytotoxic T-lymphocyte protein 4; IFN-γ, interferon-γ; IL-12, interleukin-12; PD-1, programmed cell death protein 1; SagA, secreted antigen A.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

[1]Metaorganism Immunity Section, Laboratory of Host Immunity and Microbiome, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD 20892, USA. [2]NIAID Microbiome Program, NIAID, NIH, Bethesda, MD 20892, USA. Email: ybelkaid@niaid.nih.gov; eduard.ansaldogine@nih.gov

and in responding patients in the context of cancer immunotherapy (*2*–*6*). Together, these observations argue that immunotherapy-adjuvant microbes may converge on type 1 immunity as a central effector mechanism, whereas the upstream pathways engaged by individual microbial taxa may be context and microbe dependent.

Understanding the mechanism of action of the microbiota in improving responses to immune checkpoint therapy is key for our ability to therapeutically harness them for targeted adjuvant therapies. Optimal responses to immune checkpoint therapy are likely to involve numerous non–mutually exclusive and synergistic effects of the microbiota. For example, shared antigens between an *Enterococcus* bacteriophage and a tumor antigen have been shown to lead to commensal-specific T cells that are cross-reactive with tumor antigens after anti-PD1 treatment (*15*). Additionally, defined protective bacteria can also promote commensal-specific adaptive immune responses in the context of immunotherapy (*3*, *6*). Whether these commensal-specific responses play an active role in the antitumor effects or whether they are simply a by-product of the immunostimulatory activity of these bacteria remains to be determined. Finally, translocation of commensals to the tumor bed, which can be enhanced owing to barrier disruption caused by immunotherapy, has also been proposed as a potential antitumor mechanism (*2*).

The study of Griffin *et al.* opens an avenue to harness endogenous adjuvants to fight cancer by designing targeted therapeutics that recapitulate the effects of the microbiota. This study also further illustrates the need to move away from "needle-in-the-haystack" single microbes as causal agents toward the identification of druggable canonical pathways and molecular determinants. ■

REFERENCES AND NOTES

1. E. Ansaldo, T. K. Farley, Y. Belkaid, *Annu. Rev. Immunol.* **39**, 449 (2021).
2. G. D. Sepich-Poore *et al.*, *Science* **371**, eabc4552 (2021).
3. M. Vétizou *et al.*, *Science* **350**, 1079 (2015).
4. A. Sivan *et al.*, *Science* **350**, 1084 (2015).
5. V. Matson *et al.*, *Science* **359**, 104 (2018).
6. B. Routy *et al.*, *Science* **359**, 91 (2018).
7. L. F. Mager *et al.*, *Science* **369**, 1481 (2020).
8. M. E. Griffin *et al.*, *Science* **373**, 1040 (2021).
9. D. Davar *et al.*, *Science* **371**, 595 (2021).
10. E. N. Baruch *et al.*, *Science* **371**, 602 (2021).
11. C. Iwamura *et al.*, *Blood* **129**, 171 (2017).
12. D. Kim *et al.*, *Nat. Med.* **22**, 524 (2016).
13. R. Caruso, N. Warner, N. Inohara, G. Núñez, *Immunity* **41**, 898 (2014).
14. B. Priem *et al.*, *Cell* **183**, 786 (2020).
15. A. Fluckiger *et al.*, *Science* **369**, 936 (2020).

ACKNOWLEDGMENTS

The authors are supported by the Division of Intramural Research of the National Institute of Allergy and Infectious Diseases (NIAID).

10.1126/science.abl3656

IMMUNOLOGY

Immune imprinting in utero

Mild maternal infection causes tissue-specific epigenetic imprinting in utero

By **Mohammed Amir**[1] *and* **Melody Y. Zeng**[1,2]

Increasing evidence suggests that immune system development begins in utero and is heavily influenced by the maternal immune status during gestation (*1*). Pregnancy is associated with suppression of the maternal immune system to promote the growth of the allogeneic fetus, albeit at the cost of increased susceptibility to infection (*2*). Mild maternal infections—such as mild urinary tract, skin, and respiratory tract infections—are often self-resolved and underdiagnosed. It remains poorly understood whether transient maternal inflammation induced by a mild infection leaves an immunologic scar in offspring. On page 982 of this issue, Lim *et al.* (*3*) tackle this question in mice and find that the cytokine interleukin-6 (IL-6), induced by a mild maternal infection, causes epigenetic imprinting of fetal intestine that has a long-term impact on immune responses to enteric pathogens and tissue inflammation in adulthood.

"...immune imprinting in utero... connects immune perturbations during gestation and immune predisposition in adulthood."

Recent studies increasingly depict a more active and mature fetal immune landscape, implying that fetal cells are readily susceptible to maternal immune perturbations. For example, midgestation human fetal microglia (brain-resident macrophages) exhibit phagocytic and microbial sensing properties, which may confer protection against pathogens but increase the vulnerability of the fetal central nervous system to environmental perturbations (*4*). Likewise, human fetal mast cells can be sensitized by maternal immunoglobulin E (IgE) and increase the likelihood of developing skin and airway inflammation upon allergen encounter in newborns (*5*). Microbial signals during gestation are a major driver of the early development of fetal immune cells. Microbial DNAs in fetal tissues and live microbes that may activate fetal intestinal T cells and induce immune memory before birth were detected in human fetal intestines (*1*, *6*).

Lim *et al.* report tissue-specific enrichment of T helper 17 (T_H17) cells in the small intestine in the adult offspring of dams (pregnant mice) that were infected transiently during midgestation with attenuated foodborne *Yersinia pseudotuberculosis* in the intestine. The intestinal T_H17 cells in the offspring from an infected dam were reactive to commensal resident bacteria in the intestine and remained unchanged despite cross-fostering by a naïve dam, suggesting the contribution of in utero factors induced by the mild maternal infection.

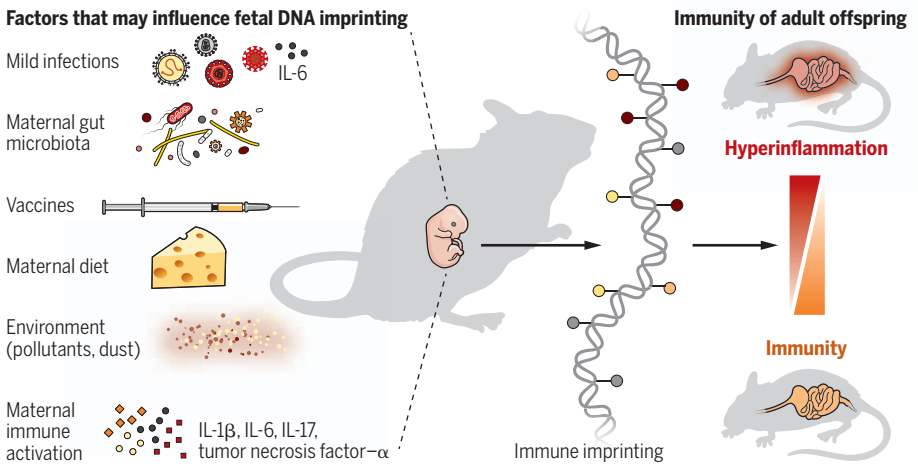
Lim *et al.* further show that the maternal infection resulted in an increase in IL-6, a cytokine that is produced mainly by macrophages during infections and promotes T_H17 cell differentiation. A single injection of IL-6 in naïve dams was sufficient to induce intestinal T_H17 cells in the adult offspring by increasing the numbers of accessible promoter regions of genes associated with epithelium development and T_H cell differentiation in fetal intestinal epithelial cells. Intriguingly, the epigenetic alterations were not coupled with changes in gene expression in fetal intestinal epithelial cells. This suggests that additional triggers are needed to induce expression of these imprinted genes but are lacking in utero. A likely trigger could be postnatal microbial colonization of the neonatal intestine.

Furthermore, Lim *et al.* demonstrate that prenatal injection of IL-6 in dams led to sustained epigenetic changes in adult intestinal stem cells, which originate from fetal intestinal epithelial cells (*7*), that favored the development of intestinal T_H17 cells. This enhanced immunity against enteric *Salmonella* infection but also increased the severity of colitis in the offspring. Additional studies are needed to assess the predisposition to inflammatory disorders beyond the intestine because T_H17 responses are associated with several autoimmune diseases, such as multiple sclerosis (*8*).

[1]Gale and Ira Drukier Institute for Children's Health, Weil Cornell Medicine, Cornell University, New York, NY, USA. [2]Department of Pediatrics, Weil Cornell Medicine, New York, NY, USA. Email: myz4001@med.cornell.edu

Shaping immune development in utero

Fetal immune development is sensitive to changes in the dam, including gut microbiota, diet, and inflammation, such as from vaccination. A mild, maternally restricted infection may lead to interleukin-6 (IL-6)–induced epigenetic imprinting in fetal intestinal epithelium and intestinal T helper 17 (T_H17) cells in adult offspring. As a result, the adult offspring have enhanced immunity against enteric pathogens but may also be more susceptible to gut inflammation.



It is still unclear why IL-6 induced by a mild maternal infection specifically targets fetal intestinal epithelial cells because other fetal cells also express the receptor, IL-6Rα. The tissue-specific imprinting by IL-6 in the fetal intestine suggests that perhaps additional signals exclusive to the fetal intestine also contribute to its heightened responsiveness to IL-6. Microbial ligands in the amniotic fluid can be swallowed by the fetus and passed through to the fetal intestine as early as midgestation (*9*); microbial stimulation of the fetal intestine may provide synergistic—but unidentified—signals in concert with maternal IL-6 to induce epigenetic modifications. Prenatal microbial exposure has emerged as a critical driver of immune development in early life (*10*-*12*). The fetal intestine, as a major site of microbial sensing and modulation, may be more receptive to in utero immune priming. Further investigation of factors in the fetal intestine that license this tissue with this distinct role in prenatal immune imprinting would elucidate potential therapeutic targets to mitigate harmful effects of maternal infections.

Maternal immune activation (MIA), often triggered by infection, increases the risk of neurodevelopmental defects in offspring. IL-6–induced maternal T_H17 cells in MIA in mice are associated with abnormal cortical changes in the fetal brain and autistic-like behaviors in offspring (*13*). Impairment in social behaviors, however, was not found in the offspring of IL-6–injected dams by Lim *et al.* The intensity and duration of the T_H17 cell response as well as the specificities of T_H17 cells may dictate the neurodevelopmental impact in the fetal brain. A mild maternal infection likely leads to a transient fluctuation of innate cytokines, such as IL-6, without eliciting a robust adaptive T cell response in the dam.

Another intriguing aspect of the study by Lim *et al.* is the elucidation of tissue-specific immune imprinting in utero as a mechanism that connects immune perturbations during gestation and immune predisposition in adulthood. The immune landscape at the maternal-fetal interface can be shaped by other factors, including the maternal gut microbiota, vaccination, and dietary changes (see the figure), all of which may affect IL-6 signaling to the fetal intestine. The past few decades have seen a marked increase in the incidence of inflammatory disorders in children, including asthma, allergies, and behavioral deficits driven in part by neuroinflammation (*14*). Future work should address whether and how immune imprinting in utero may underlie the predisposition to inflammatory disorders. ■

REFERENCES AND NOTES

1. A. Mishra *et al.*, *Cell* **184**, 3394 (2021).
2. G. Mor, P. Aldo, A. B. Alvero, *Nat. Rev. Immunol.* **17**, 469 (2017).
3. A. I. Lim *et al.*, *Science* **373**, eabf3002 (2021).
4. L. Kracht *et al.*, *Science* **369**, 530 (2020).
5. R. Msallam *et al.*, *Science* **370**, 941 (2020).
6. N. Li *et al.*, *Nat. Immunol.* **20**, 301 (2019).
7. J. Guiu *et al.*, *Nature* **570**, 107 (2019).
8. T. Moser, K. Akgün, U. Proschmann, J. Sellner, T. Ziemssen, *Autoimmun. Rev.* **19**, 102647 (2020).
9. A. N. Ardissone *et al.*, *PLOS ONE* **9**, e90784 (2014).
10. M. L. Conrad *et al.*, *J. Exp. Med.* **206**, 2869 (2009).
11. M. J. Ege *et al.*, *J. Allergy Clin. Immunol.* **122**, 407, 412.e1 (2008).
12. M. Gomez de Agüero *et al.*, *Science* **351**, 1296 (2016).
13. G. B. Choi *et al.*, *Science* **351**, 933 (2016).
14. R. X. Y. Chua *et al.*, *Front. Neurol.* **11**, 603571 (2021).

ACKNOWLEDGMENTS

M.Y.Z. is supported by NIH career transition grant K01 DK114376.

10.1126/science.abl3631

VIEWPOINT: COVID-19

The animal origin of SARS-CoV-2

Trading of animals susceptible to bat coronaviruses is the likely cause of the COVID-19 pandemic

By **Spyros Lytras[1], Wei Xia[2], Joseph Hughes[1], Xiaowei Jiang[3], David L. Robertson[1]**

Although first detected in December 2019, COVID-19 was inferred to be present in Hubei province, China, for about a month before (*1*). Where did this new human disease come from? To understand the origin of the COVID-19 pandemic, it is necessary to go back to 2002. At that time a novel respiratory coronavirus appeared in Foshan, Guangdong province, China, and spread to 29 countries (*2*). Altogether ~8000 people were infected with severe acute respiratory syndrome coronavirus (SARS-CoV) before public health measures controlled its spread in 2003. The zoonotic origin of SARS-CoV was subsequently linked to live animals available at markets. Further sporadic spillover events of SARS-CoV from animals took place in Guangzhou, Guangdong, and some researchers working with cultured virus were infected in laboratory accidents (*3*), but ultimately SARS-CoV was removed from the human population. Trading of susceptible host animals is an important common theme in the emergence of SARS and COVID-19.

Three years after the SARS epidemic began, investigations revealed that horseshoe bats (*Rhinolophus*) in China were harboring related coronaviruses (*4*). These collectively form the species SARS-related coronavirus (SARSr-CoV), which comprises the *Sarbecovirus* subgenus of the *Betacoronavirus* genus. It was inferred that a sarbecovirus circulating in horseshoe bats seeded the progenitor of SARS-CoV in an intermediate animal host, most probably civet

[1]Medical Research Council–University of Glasgow Centre for Virus Research, Glasgow, UK. [2]National School of Agricultural Institution and Development, South China Agricultural University, Guangzhou, China. [3]Department of Biological Sciences, Xi'an Jiaotong–Liverpool University, Suzhou, China. Email: xiaowei.jiang@xjtlu.edu.cn; david.l.robertson@glasgow.ac.uk

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

cats (*3*). Although other possible intermediate hosts for SARS-CoV were identified, in particular raccoon dogs and badgers (for sale with civet cats in animal markets), it is a population of civet cats within markets that appear to have acted as the conduits of transmission to humans from the horseshoe bat reservoir of SARS-CoV, rather than civet cats being a long-term reservoir host species. Presumably a captive civet cat initially became infected by direct contact with bats—e.g., as a result of bats foraging in farms or markets—or was infected prior to capture. Following the SARS epidemic, further surveillance revealed the immediate threat posed by sarbecoviruses from horseshoe bats. Despite this clear warning, another member of the SARSr-CoV species, SARS-CoV-2, emerged in 2019 that spread with unprecedented efficiency among humans. There has been speculation that the Wuhan Institute of Virology (WIV) in Hubei was the source of the pandemic because no SARS-CoV-2 intermediate host has been identified to date and owing to the WIV's geographic location.

SARS-CoV-2 first emerged in Wuhan city, which is >1500 km from the closest known naturally occurring sarbecovirus collected from horseshoe bats in Yunnan province, leading to an apparent puzzle: How did SARS-CoV-2 arrive in Wuhan? Since its emergence, sampling has revealed that coronaviruses genetically close to SARS-CoV-2 are circulating in horseshoe bats, which are dispersed widely from East to West China, and in Southeast Asia and Japan (*5*). The wide geographic ranges of the potential reservoir hosts—for example, intermediate (*R. affinis*) or least (*R. pusillus*) horseshoe bat species, which are known to be infected with sarbecoviruses—indicate that the singular focus on Yunnan is misplaced (*5*). Confirming this assertion, the evolutionarily closest bat sarbecoviruses are estimated to share a common ancestor with SARS-CoV-2 at least 40 years ago (*5*), showing that these Yunnan-collected viruses are highly divergent from the SARS-CoV-2 progenitor. The first of these viruses reported by WIV, RaTG13 (*6*), is certainly too divergent to be the SARS-CoV-2 progenitor, providing key genetic evidence that weakens the "lab-leak" notion. Additionally, three other sarbecoviruses collected in Yunnan independently of the WIV are now the closest bat coronaviruses to SARS-CoV-2 that have been identified: RmYN02, RpYN06, and PrC31 (see the figure).

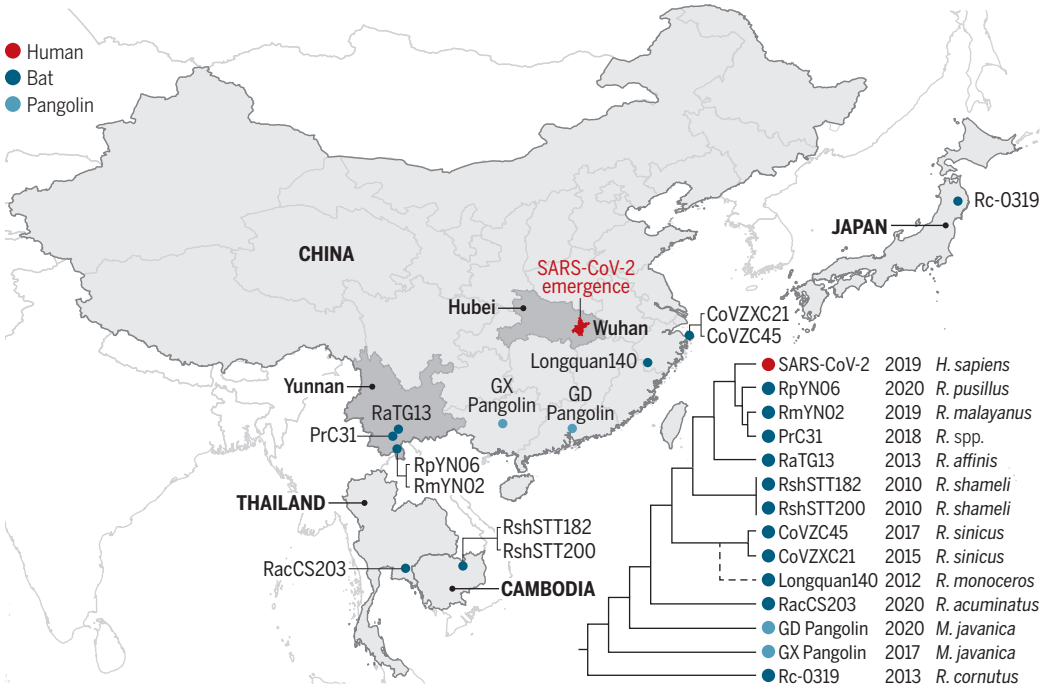
So, how did SARS-CoV-2 get into humans? Although it is possible that a virus spillover occurred through direct horseshoe bat–to–human contact, a known risk for SARSr-CoVs (*7*), the first detected SARS-CoV-2 cases in December 2019 are associated with Wuhan wet markets (*8*). This is consistent with multiple animal-market–associated spillover events in November and December (*9*). It is currently not possible to be certain of the animal source of SARS-CoV-2, but it is notable that live animals, including civet cats, foxes, minks, and raccoon dogs, all susceptible to sarbecoviruses, were for sale in Wuhan markets, including the Huanan market (identified as an epicenter of the outbreak in Wuhan) throughout 2019 (*10*). Many of these animals are farmed for their fur at large scale and then sold to animal markets (*11*). Some of these farmed species (American minks, red foxes, and raccoon dogs) were sold alive for food by Wuhan animal sellers, as was trapped wildlife (including raccoon dogs and badgers), although no bat species were for sale (*10*). Together, this suggests a central role for SARSr-CoV–susceptible live intermediate host animals as the primary source of the SARS-CoV-2 progenitor that humans were exposed to, as was the case with the origin of SARS.

If these routes of transmission to humans are in place, why is emergence so rare that only two major outbreaks have occurred in the last two decades? Spillover events are not so unusual in locations where more frequent human-animal contacts take place. This is indicated by serology studies showing evidence for SARSr-CoV–specific antibodies in people living in rural locations (*12*), and even higher rates recorded in people living near bat caves (*7*). Spillover risk will increase with human encroachment into rural areas, resulting from new travel networks around and between urban areas. When a novel virus is then exposed to a densely packed human population, such as in Wuhan city, these spillover events have a much higher chance of resulting in substantial onward spread (*1*).

One particular ecological event in China that severely disrupted meat trade, and thereby contributed to increased wildlife–human contacts, was the shortage of pork products in 2019. This was a direct consequence of the African swine fever virus (ASFV) pandemic (*11*), which led to ~150 million pigs being culled in China, resulting in a pork supply reduction of ~11.5 million metric tons in 2019. Although production of other meat, such as poultry, beef,

Sarbecoviruses closely related to SARS-CoV-2

Coronaviruses that are evolutionarily closest to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been sampled in China, Cambodia, Japan, and Thailand (*5*). The phylogenetic tree, inferred from a genomic region minimized for recombination (*5*), shows sarbecoviruses closely related to SARS-CoV-2. Host species for each virus, horseshoe bat (*Rhinolophus*), human (*Homo sapiens*), and pangolin (*Manis javanica*) and the year of sample collection are shown in the key. Longquan140 is inferred from another genomic region (*5*) (dashed line). See supplementary table S1 for more details.



GRAPHIC: K. FRANKLIN/*SCIENCE*

and fish products, moderately increased and China imported more of these products from international markets to mitigate the shortfall, this supply only covered a fraction of the ASFV-associated pork losses. Consequently, pork prices hit a record high in November 2019, with the wholesale price increasing ~2.3 times compared with the previous year. Moreover, pig production has been relocating from Southern to Northern China since 2016. This, coupled with tight restrictions on the movement of live pigs and pork products to mitigate the ASFV pandemic, reduced the availability of pork in the Eastern and Southern provinces, resulting in much steeper price increases in these regions. In response, food consumers and producers may have resorted to alternative meats, including farmed or captured wildlife, especially in Southern China where wildlife is traditionally consumed (*11*). The resulting increased trade of susceptible farmed animals and wildlife could have brought humans into more frequent contact with meat products and animals infected with zoonotic pathogens, including SARSr-CoVs.

There are controversial reports of human SARS-CoV-2 cases in China being traced back to contact with imported frozen foods and SARS-CoV-2 apparently identified from frozen food, packaging, and storage surfaces (*13*). In an effort to prevent ASFV spread through live pig transportation routes, supply through the cold chain has been encouraged by the Chinese government since October 2018, with stronger support since September 2019 in the form of waiving freeway toll fees for frozen pork. The large demand for pork meat facilitated the use of cold-chain transport for all meat types, in particular from places with lower prices to those with higher prices, legally (or illegally), potentially also including transport of species susceptible to SARSr-CoV infection. The World Health Organization (WHO) Origins Report (*8*) recorded carcasses of wildlife, particularly badgers, left behind in freezers at the Huanan market, as well as their sale as frozen goods in late December 2019. It is likely that this wildlife had been trapped or farmed elsewhere and sold to Wuhan markets through the cold chain. Exposures could also potentially occur through feeding of coronavirus-infected carcasses to live animals either in transport or at markets.

The emergence of SARS-CoV-2 has properties that are consistent with a natural spillover (*9*). Although carriage from a bat cave of a sarbecovirus close enough to SARS-CoV-2 to be the progenitor as a research sample to the WIV is theoretically possible, such a scenario would be extremely unlikely relative to the scale of human-susceptible animal contacts routinely taking place in animal trading. Alternatively, bat guano (feces) is collected for use as fertilizer, again on a much larger scale than irregular research visits to bat caves, consistent with rare but ongoing SARSr-CoV transmissions to humans in rural areas (*7, 12*).

Overall, SARSr-CoV animal-to-human transmission associated with infected live animals is the most likely cause of the COVID-19 pandemic. However, the massive scale of cold-chain supply, particularly following disruption to the meat industry in China caused by ASFV-associated culling, suggests that frozen susceptible-animal carcasses, either for human or animal consumption, should not be discounted as playing a role in the emergence of SARS-CoV-2. This will especially be the case if the progenitor population of SARS-CoV-2 is found further away from Wuhan, because live-animal trafficking is much more likely to involve more proximal locations to the city, e.g., the prefectures of Hubei province. Serology, sampling and interviewing of the individuals (e.g., trappers, traders, and farmers) connected to the sources of wildlife sold in the Wuhan markets in October and November 2019 would be a sensible next step in future investigations.



Horseshoe bats, such as this greater horseshoe bat (*Rhinolophus ferrumequinum*), can be a reservoir of coronaviruses.

Once in the human population, SARS-CoV-2 has spread surprisingly rapidly for a new human pathogen. Contrary to classical expectations for a host species jump, SARS-CoV-2 is highly capable of human transmission, including frequent asymptomatic transmission and amplification through superspreader events. This initial "success," at least prior to the emergence of variants of concern, is unlikely to be due to early adaptation to humans but rather can be attributed to the relatively generalist nature of SARS-CoV-2 (*14*), evidenced by frequent transmission to mammals: minks, cats, and others. Worryingly, recent experimental evidence has found that the pangolin-derived sarbecoviruses (presumably acquired from exposure to horseshoe bats or other infected animals after illegal trafficking into China) can also infect human cells and have spike proteins that are even better at facilitating entry into human cells than that of SARS-CoV-2 (*15*). Collectively this points to a further risk of spillover that extends to the more divergent members of the lineage that SARS-CoV-2 emerged from and implies frequent spillovers from bats to other susceptible wildlife.

Humans are now the dominant SARS-CoV-2 host species. The danger is that SARS-CoV-2 could spread from humans to other animal species, termed reverse zoonosis, as is suspected for white-tailed deer in the United States. The promiscuous infection of various host species by the sarbecoviruses means that future spillovers of SARSr-CoVs from wildlife are very likely, and current vaccines may not be protective against novel variants. The sampling intensity of sarbecoviruses needs to be urgently increased to gain a better understanding of this spillover risk. The recent finding of sarbecoviruses, not dissimilar to SARS-CoV-2, dispersed in Southeast Asia emphasizes the urgency of monitoring coronavirus diversity. Humanity must work together beyond country borders to amplify surveillance for coronaviruses at the human–animal interface to minimize the threat of both established and evolving variants evading vaccines and to stop future spillover events. ■

REFERENCES AND NOTES

1. J. Pekar *et al.*, *Science* **372**, 412 (2021).
2. T. G. Ksiazek *et al.*, *N. Engl. J. Med.* **348**, 1953 (2003).
3. H.-D. Song *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2430 (2005).
4. W. Li *et al.*, *Science* **310**, 676 (2005).
5. S. Lytras *et al.*, *bioRxiv* 10.1101/2021.01.22.427830 (2021).
6. P. Zhou *et al.*, *Nature* **579**, 270 (2020).
7. N. Wang *et al.*, *Virol. Sin.* **33**, 104 (2018).
8. WHO-convened global study of origins of SARS-CoV-2: China Part; www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part (2021).
9. E. C. Holmes *et al.*, *Cell* 10.1016/j.cell.2021.08.017 (2021).
10. X. Xiao *et al.*, *Sci. Rep.* **11**, 11898 (2021).
11. W. Xia *et al.*, *Preprints*, 10.20944/preprints202102.0590.v1 (2021).
12. H. Li *et al.*, *Biosaf Health* **1**, 84 (2019).
13. J. Han *et al.*, *Environ. Chem. Lett.* **19**, 5 (2020).
14. C. Conceicao *et al.*, *PLOS Biol.* **18**, e3001016 (2020).
15. S. J. Dicken *et al.*, *bioRxiv* 10.1101/2021.03.22.436468 (2021).

ACKNOWLEDGMENTS

We thank the researchers who have shared genome data openly via GenBank or GISAID. D.L.R. and J.H. are funded by the Medical Research Council (MRC) (MC_UU_1201412) and D.L.R. by the Wellcome Trust (220977/Z/20/Z). S.L. is funded by an MRC studentship.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/968/suppl/DC1

Published online 17 August 2021
10.1126/science.abh0117

PHOTO: DIETMAR NILL/MINDEN PICTURES



BOOKS *et al.*

Visitors experience Chrystel Lebas's immersive installation, *Regarding Forests*.

PHOTO: *REGARDING FORESTS*, CHRYSTEL LEBAS 2021. *TRANQUILLITY*, WELLCOME COLLECTION. STEVEN POCOCK, 2021/CC-BY-NC

EXHIBITION

The spectrum of happiness

A pair of exhibitions challenge visitors to embrace a more encompassing view of contentment

By **Vasilisa Skvortsova**

On a recent morning, while navigating toward the Wellcome Collection to see its latest offerings, *Joy* and *Tranquillity*—two exhibitions featured as part of the museum's "On Happiness" season—I found myself wondering what could be said about happiness that had not already been articulated. From ancient texts to apps that ask users to rate their hourly happiness level (*1*) to the multibillion-dollar wellness industry, the hunt for happiness seems to be universal. And yet it remains as elusive as it is mundane. Happily, the exhibitions offer an unconventional perspective on the types of things that make us feel good.

Tranquillity
Laurie Britton Newell and George Vasey, curators
Wellcome Collection, London, UK,
15 July 2021–9 January 2022

Joy
Laurie Britton Newell and George Vasey, curators
Wellcome Collection, London, UK,
15 July 2021–27 February 2022

The idea for these exhibitions originated in 2018, but a number of events, including the COVID-19 pandemic, pushed curators Laurie Britton Newell and George Vasey to "urgently reframe initial questions" and shift the focus toward a more "entangled version of happiness" as a way to "navigate moments of adversity." One of the goals of the exhibitions, according to Vasey, is to engage with visitors and to bring catharsis, as summarized by a note written by Octavia Butler to herself: "Make people feel! Feel! Feel!" (The diary in which this passage appears is on display in *Tranquillity*.)

Upon entering, visitors are immersed in a dimly lit space with a mesmerizing crystal in the center, an installation created by Scottish Indian artist Jasleen Kaur that is meant to evoke a yoga studio. But relaxation is not the point of this space. It is meant instead to question the selfish pursuit of personal well-being at the expense of others and the environment.

From a cosmos-inspired folding almanac used to diagnose medical conditions in the 15th century to British photographer Toby Glanville's images that capture the power of communal gardening, the exhibition documents the toolkit that humanity has built over the centuries to balance its passions. None of these mediums is as powerful as an encounter with nature, however. With the goal of evoking this singular experience, French photographer Chrystel Lebas's multisensory installation, *Regarding Forests*, features large-format photographs of some of the oldest trees in the world, earthy scents, and a soundscape of rainforest fauna, recreating the Japanese practice known as "shinrin-yoku" (forest bathing).

"Pleasure comes in repeated cycles, with different phases of wanting, liking, and satisfaction or satiety," notes Morten Kringelbach, professor of neuroscience at Aarhus University, Denmark, and one of the exhibition's advisers in one of the exhibits. Likewise, our emotions do not arise in isolation but influence, and are influenced by, our proximity to others. The *Tranquillity* exhibition thus prepares visitors for an encounter with the other end of the emotional spectrum, *Joy*, an exhibition on display in the museum's second gallery.

Visitors to *Joy* are greeted with a simulated dance floor of ecstatic figures in a room painted in vibrant yellow created by Harold Offeh. Dance is known for its therapeutic and pain-healing properties and for its ability to translate mental states into body movements.

The exhibition also highlights drawings by the 11th-century Islamic physician Abu Ali Ibn Sina (Avicenna) that foreshadow our growing appreciation for the role played by brain-gut interactions in depression (*2*) and the beneficial role family and friendship have on our life expectancy, as observed by British anthropologist Robin Dunbar (*3*). Before one can start nervously calculating whether her friend circle is large enough to see her through to old age, however, she is likely to be distracted by colorful placards and the festive music playing in the next room. Here, visitors are confronted with the following question: Can protest be pleasurable? In their works, artists Amalia Pica and Joshua Virasami argue that it can indeed be, highlighting examples of "pleasure activism" (*4*) that reveal how striving for social justice could be a way to rewire perturbed social connections.

The somewhat eclectic agglomeration of historical objects, scientific derivations, and contemporary art objects that are brought together in this pair of exhibitions hints at the difficulty of deconstructing a complex concept like happiness—a term that the curators intentionally avoid using. But they accurately reflect that there is no one-size-fits-all formula for happiness. *Joy* and *Tranquillity* represent a good starting point to increase our collective emotional literacy. ■

The reviewer is at the Max Planck UCL Centre for Computational Psychiatry and Ageing Research and the UCL Queen Square Institute of Neurology, University College London (UCL), London WC1B 5EH, UK.
Email: v.skvortsova@ucl.ac.uk

REFERENCES AND NOTES

1. The Happiness Project, https://thehappinessproject.app/.
2. M. Valles-Colomer *et al.*, *Nat. Microbiol.* **4**, 623 (2019).
3. R. Dunbar, *Friends: Understanding the Power of Our Most Important Relationships* (Little, Brown Book Group, 2021).
4. A. M. Brown, *Pleasure Activism: The Politics of Feeling Good* (AK Press, 2019).

10.1126/science.abk3121

HISTORY OF MEDICINE

The unsung players of epidemiology

A new history probes often-overlooked contributions to the study of infectious disease

By **Suman Seth**

Maladies of Empire: How Colonialism, Slavery, and War Transformed Medicine
Jim Downs
Belknap Press, 2021. 272 pp.



When the British ship the *Eclair* arrived at Boa Vista, an island off the coast of West Africa, in August 1845, fever was raging through its crew. A local surgeon declared that the illness was not contagious, and authorities permitted sick sailors to be housed on a nearby island. The *Eclair* left after 3 weeks, the fever having ravaged its men. Within a few days, the soldiers who had guarded the crew began to fall ill. Manoel Antonio Alves, another local soldier, described in the records as a "negro," was ordered to retrieve the bodies of those who died. Roughly 6 months later, Alves and other soldiers recounted their experiences to British naval surgeon James McWilliam, whose *Report on the Fever at Boa Vista* would be published in 1847.

In many medical histories, McWilliam's would be the only name referenced in an analysis of this outbreak. In Jim Downs's account of the roots of modern epidemiology, however, it is Alves and individuals like him—soldiers as well as washerwomen, hospital attendants, the enslaved, Muslim pilgrims, and others—who take center stage. "[M]any modern epidemiological practices grew in part out of observing, treating, and preventing disease among captive populations produced by colonialism, slavery, and war," Downs writes. For those of us looking warily toward future epidemics, this book draws our attention to oft-forgotten sources of medical knowledge.

The formal study of epidemics goes back at least to Hippocrates, and as such, the dates bounding Downs's study—from 1756, the year of the so-called Black Hole of Calcutta jail incident, to 1866, the year that saw the end of the third major cholera outbreak in the United States—feel somewhat arbitrary. That said, the crowded jail cell in which 123 of 146 British soldiers died in the Indian heat in 1756 is an excellent example of the type of event from which 18th-century knowledge concerning the value of adequate ventilation would emerge.

The reviewer is at the Department of Science and Technology Studies, Cornell University, Ithaca, NY 14853, USA. Email: ss536@cornell.edu

In chapter two, the book turns to Malta in the 1830s, where the fact that laundresses often failed to contract the plague after handling soiled linens was used to argue against the need for quarantines. A substantial middle chapter makes the case for seeing Florence Nightingale as an "unrecognized epidemiologist." Because Nightingale's contributions to medical statistics are well known to historians, this claim seems intended to correct a popular rather than a scholarly misconception.



Laundresses who handled soiled linens often failed to contract the plague.

The influence of Nightingale and other sanitarian reformers can be clearly seen in the foundation of the United States Sanitary Commission. Downs's chapter on the commission stresses the racialist logics the group brought to its epidemiological work, with its persistent emphasis on the differences in susceptibilities and immunities to disease between white and Black troops. Somewhat oddly, this is contrasted with British attitudes, which are depicted as much more concerned with environmental causes than racial difference. This opposition, however, is artificial: Sanitarians believed that many diseases were caused by external conditions, yet such conditions also had different effects on bodies that varied according to gender, race, class, or age.

Race was a central category of analysis in the four large and enormously influential statistical reports (published between 1838 and 1841) on the health of troops throughout the British Empire, and racial statistics were used to justify pulling white troops out of dangerous locations in the West Indies and Western Africa and replacing them with soldiers of African descent. Rather than being an outlier, the US was part of a much broader trend at the time toward the racialization of many fields, medicine included.

More convincing is a chapter on smallpox vaccination during the US Civil War. Running low on material with which to protect their troops, the Confederate army sought to produce supplies of their own by harvesting lymph from the sores of the vaccinated. Rejecting the wearied bodies of soldiers as sources, doctors turned instead to enslaved children, who, Downs writes, "were most likely used as the primary source of vaccine matter in the Civil War South." Here we see Downs's thesis in its starkest form, with the bodies of the most vulnerable used to avoid epidemics among those with more power.

Historians of medicine, race, and colonialism will take issue with some of the claims made in Downs's chapters. His overall thesis is an important one, however, and the book deserves to be read, particularly now. Few will question the salvational power that epidemiology will likely have in the years to come. But we should keep in mind the contributions, willing and otherwise, of those on whom our knowledge relies. ■

10.1126/science.abl3828

SCIENCE & RACE



PODCAST

Breathing Race into the Machine: The Surprising Career of the Spirometer from Plantation to Genetics
Lundy Braun
University of Minnesota Press, 2014. 304 pp.

The spirometer is, at first glance, an innocuous instrument used to determine an individual's lung capacity. However, its measurements are often "race corrected" to account for a supposed difference between white and Black patients, a practice that can influence whether an individual qualifies for employment or meets disability criteria. This week on the *Science* podcast, Lundy Braun elaborates on the shaky science of the spirometer and explores the broader implications of race correction in medicine.
https://scim.ag/3CY9Lq9

10.1126/science.abl8016

IMAGE: ISTOCK.COM/CLU



Feral equids, such as these donkeys in the US Mojave desert, affect desert ecosystems by digging wells to access water.

PHOTO: MICHAEL ALFUSO

Edited by **Jennifer Sills**

Feral equids' varied effects on ecosystems

In their Report "Equids engineer desert water availability" (30 April, p. 491), E. J. Lundgren and coauthors present observations of feral burros digging shallow wells and suggest that these wells may benefit native wildlife and ecosystems by increasing the number of water sites. The authors consider feral equids (horses and burros) to be beneficial but fail to acknowledge substantial research establishing the many negative effects that feral equids have on arid ecosystems. The positive view presented by Lundgren *et al.* is troubling because growing feral equid populations in the western United States are a serious concern for natural resource managers (*1*).

Feral equids compete with native wildlife for food and water and displace native ungulates from water (*2*–*5*). Their presence has been associated with decreased diversity of native species at water, and native species have been observed to visit water less frequently and spend less time at water when equids are present (*6*). Reduced access to water is particularly detrimental in arid landscapes, especially during drought.

The addition of a small number of shallow wells does not benefit native wildlife when their overall access to water is reduced by equids. Lundgren *et al.* present no data to demonstrate an overall benefit to wildlife. We urge caution in speculating that the addition of a few localized wells benefits native species. Feral equids have also altered plant composition, reduced plant diversity and cover, and increased soil compaction and erosion (*7*–*9*), thereby putting native habitats at risk. This, together with poorly controlled equid population growth, results in degraded habitats with decreased carrying capacities. Effective management of feral equids will benefit from the public's accurate understanding of their overall effects on ecosystems.

Esther S. Rubin[1]*, Dave Conrad[2], Andrew S. Jones[1], John J. Hervert[1]
[1]Arizona Game and Fish Department, Phoenix, AZ 85086, USA. [2]Phoenix, AZ 85086, USA.
*Corresponding author. Email: erubin@azgfd.gov

REFERENCES AND NOTES

1. National Research Council, *Using Science to Improve the BLM Wild Horse and Burro Program: A Way Forward* (The National Academies Press, Washington, DC, 2013).
2. S. Ostermann-Kelm, E. R. Atwill, E. S. Rubin, M. C. Jorgensen, W. M. Boyce, *J. Mammal.* **89**, 459 (2008).
3. N. D. Perry, P. Morey, G. San Miguel, *Southwest. Natur.* **60**, 390 (2015).
4. A. M. J. Gooch *et al.*, *J. Arid Environ.* **138**, 38 (2017).
5. L. K. Hall, R. T. Larsen, R. N. Knight, B. R. McMillan, *Ecosphere* **9**, e02096 (2018).
6. L. K. Hall *et al.*, *J. Arid Environ.* **127**, 100 (2015).
7. E. A. Beever, L. Huntsinger, S. L. Petersen, *Biol. Conserv.* **226**, 321 (2018).
8. M. M. Kaweck, J. P. Severson, K. L. Launchbaugh, *Rangelands* **40**, 45 (2018).
9. K. W. Davies, C. S. Boyd, *BioScience* **69**, 558 (2019).

10.1126/science.abl5863

Response

Rubin *et al.* rightly point out that feral equids have diverse effects on ecosystems, including exerting strong herbivory pressure and dominating water sources. These effects are common to all large herbivores [e.g., (*1*–*3*)]. If we did not already know that feral equids were introduced, would we be able to tell from their actual effects? Just as elephants can dominate water sources in Africa (*4*), most animal communities show dominance hierarchies at limiting resources [e.g., (*5*)]. Preliminary evidence indicates that this is true for feral donkeys, but only where cougars are absent (*6*).

The effects of organisms depend on the ecological context of the population. Feral equids can be the primary prey of cougars (*7*). Yet predation, as well as the near-ubiquitous persecution of predators, has been overlooked in nearly all research regarding the effects of feral equids—including in the studies cited by Rubin *et al.* Much as the return of wolves to Yellowstone National Park changed the behavior and thus the effects of elk (*8*), the effects of feral equids would likely change if we protected their predators.

The effects of organisms are also scale-dependent and diverse, and they may appear contradictory. For example, beaver herbivory can deforest ecologically important riparian woodlands, yet beavers can also create ecologically important wetlands (*9*). Elephant disturbance can create

essential habitat for small vertebrates at the patch scale but appear to suppress their diversity and abundance at broader scales (*3*, *10*). Likewise, while feral equids can have all the effects that Rubin *et al.* describe, our Report shows that they can also dig wells up to 2 m in depth—at times providing the only water available. Feral donkey disturbance also appears important to maintaining open-water habitat at desert springs; several endemic and endangered freshwater fish populations in both North America and Australia went extinct following feral donkey eradications (*11*). Describing feral equids as either ecological heroes or as invasive pests oversimplifies complexity and moralizes ecology.

Feral equids, like other introduced species, can change ecosystems in ways that alarm us. However, our desire to fight these changes can hinder our ability to study them with openness and curiosity, and in light of Earth's dynamic history (*12*).

Erick J. Lundgren[1,2,3]*, Daniel Ramp[1], Jianguo Wu[4,5], Martin Sluk[6], Karla T. Moeller[4], Juliet C. Stromberg[4], Arian D. Wallach[1]
[1]Centre for Compassionate Conservation, Faculty of Science, University of Technology Sydney, Ultimo, NSW, Australia. [2]Center for Biodiversity Dynamics in a Changing World, Department of Biology, Aarhus University, Aarhus, Denmark. [3]Section for Ecoinformatics and Biodiversity, Department of Biology, Aarhus University, Aarhus C, Denmark. [4]School of Life Sciences, Arizona State University, AZ 85281, USA. [5]School of Sustainability, Arizona State University, AZ 85281, USA. [6]Roger Williams Park Museum of Natural History, Providence, RI 02907, USA.
*Corresponding author.
Email: erick.lundgren@gmail.com

REFERENCES AND NOTES

1. V. J. Dodge, V. T. Eviner, J. H. Cushman, *Ecol. Evol.* **10**, 10858 (2020).
2. J. R. Goheen *et al.*, *PLOS One* **8**, e55192 (2013).
3. R. M. Pringle, T. P. Young, D. I. Rubenstein, D. J. McCauley, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 193 (2007).
4. M. Valeix, S. Chamaillé-Jammes, H. Fritz, *Oecologia* **153**, 739 (2007).
5. S. Rabinowicz *et al.*, *PLOS One* **15**, e0244299 (2021).
6. E. J. Lundgren *et al.*, *bioRxiv*, 10.1101/2021.04.13.439662 (2021).
7. A. M. Andreasen, K. M. Stewart, W. S. Longland, J. P. Beckmann, *J. Wildl. Manag.* **85**, 1104 (2021).
8. D. Fortin *et al.*, *Ecology* **86**, 1320 (2005).
9. P. P. Gibson, J. D. Olden, *Aquat. Conserv.* **24**, 391 (2014).
10. R. M. Pringle, *Ecology* **89**, 26 (2008).
11. A. Kodric-Brown, J. H. Brown, *Front. Ecol. Environ.* **5**, 549 (2007).
12. E. J. Lundgren *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 7871 (2020).

10.1126/science.abl7466

Academic bullying: How to be an ally

Academic bullying and harassment are all too common, and institutional reactions are often inadequate (*1*, *2*). Although reports often focus on what the targets of bullying can do to protect themselves [e.g., (*3*)], it is important to acknowledge that all members of the scientific community can and must address academic bullying. If witnesses to harassment, members of investigative committees, journal editors, other gatekeepers, and individuals at every level of the scientific community all leverage their knowledge and power to combat bullying, we can create a safer and more civil scientific environment.

Witnesses of academic bullying can report the incident to trustworthy resources at the relevant institution and funding agency. For example, in the United States, one could report unacceptable behavior to a university ombudsman or submit a bullying report to the US National Institutes of Health (NIH) (*4*). The NIH has removed more than 70 lab leaders from their grants (*5*), demonstrating that such reports by witnesses can be effective. The UK Wellcome Trust also invites such reports (*6*) and has withdrawn funding from scientists accordingly [e.g., (*7*)]. Actions by funding agencies can prevent decisions that just pass the harasser (*8*) to a different environment where they can focus on new targets.

Members of institutional academic bullying investigation committees can take steps to get unbiased and comprehensive reports of the bullying behavior. To do so, they should ask former lab members about the perpetrator, given that their feedback is likely to be more accurate, honest, and informative than that of current lab members, whose serious reservations might include fear of retaliation. Committee members can support targets during the course of investigation by putting in place a constant monitoring system to prevent retaliation against the target. They can also avoid unnecessary delays in the investigation process, while making sure not to sacrifice quality for speed. Efficient investigations show respect for the target and minimize visa issues for international students. To increase transparency, committees should make available at least one example of the outcome of a previous investigation, including discipline and punishment of the perpetrator and support for the target.

Journal editors can be open to submissions about sexual harassment, academic bullying, and imbalances in diversity (*9*), regardless of the main scope of the journal. The publication's readership, regardless of discipline, has a right to be aware of these issues. Publishing such information empowers readers to contribute to the global efforts to address these issues.

Members of gatekeeping organizations for scientific metrics (such as institution and hospital rankings) and clubs (such as national academies of sciences) can consider academic bullying in their criteria. The US National Academy of Sciences, for example, ejected members in response to validated sexual harassment allegations (*10*). In 2018, AAAS (the publisher of *Science*) established a new policy for ejecting harassers among AAAS fellows (*11*).

Individuals in all positions in the scientific community, regardless of their seniority or whether they have witnessed offensive behavior, can help address inequities, harassment, bullying, and discrimination in STEM by refusing to remain neutral about these issues. They can demand more accountability and more transparency about expectations, consequences, and reporting guidelines. They can also work to increase awareness about academic bullying and harassment by sharing studies, reports, and their own experiences.

Evidence shows that incidents of sexual harassment do not decrease simply as a result of strong policies and legal recourse (*12*). Diminishing academic harassment in science requires attention and collaborative action by all members of the scientific workforce (*8*).

Morteza Mahmoudi
Department of Radiology and Precision Health Program, Michigan State University, East Lansing, MI 48824, USA. Email: mahmou22@msu.edu

REFERENCES AND NOTES

1. T. L. Goulet, *Science* **373**, 170 (2021).
2. A. Witze, *Nature* **595**, 15 (2021).
3. M. Mahmoudi, *Nat. Hum. Behav.* **4**, 1091 (2020).
4. NIH, "NIH process for handling allegations of sexual harassment on an NIH-funded project at a recipient institution (2020); https://grants.nih.gov/grants/policy/harassment/actions-oversight/allegation-process.htm.
5. J. Kaiser, "More than 70 lab heads removed from NIH grants after harassment findings," *Science* 10.1126/science.abj9518 (2021).
6. Wellcome Trust, "Bullying and harassment policy" (2021); https://wellcome.org/grant-funding/guidance/bullying-and-harassment-policy.
7. H. Else, *Nature* **560**, 420 (2018).
8. J. Mervis, *Science* **366**, 1057 (2019).
9. M. Mahmoudi, L. Keashly, *Angewandte Chem.* **133**, 3378 (2021).
10. M. Wadman, *Science* **372**, 224 (2021).
11. M. Wadman, *Science* **361**, 1175 (2018).
12. National Academies of Sciences, Engineering, and Medicine, "Sexual harassment of women: Climate, culture, and consequences in academic sciences, engineering, and medicine" (2018).

COMPETING INTERESTS

M.M. is a co-founder and director of the Academic Parity Movement (www.paritymovement.org), a nonprofit organization dedicated to addressing academic discrimination, violence, and incivility, and receives royalties/honoraria for his published books, plenary lectures, and licensed patents.

10.1126/science.abl7492



#IfThenSheCan – The Exhibit, at Dallas's NorthPark Center, includes more than 120 ambassador statues.

PHOTO: IF/THEN

IF/THEN ambassadors find innovative ways to connect

Ambassadors are role models for middle-school girls—and each other

By **Becky Ham**

In September 2019, the AAAS IF/THEN Ambassadors program selected 125 women in science, technology, engineering, and mathematics (STEM) careers to serve as role models for middle-school girls. As part of IF/THEN, a national initiative of Lyda Hill Philanthropies to encourage more women in STEM, ambassadors were soon starring on television and on YouTube and working with girls in museums, classrooms, and Girl Scout meetings across the United States.

Some of the ambassadors' original engagement plans were upended by COVID-19, but many of the women found inspiration in a tumultuous year. "The ambassadors have been very innovative in finding ways to continue to engage middle-school girls throughout the pandemic," said Emily Therese Cloyd, director of the AAAS Center for Public Engagement with Science and Technology, who heads up the ambassadors program.

As part of the program, ambassadors could apply for $10,000 She Can Change the World grants to develop a wide variety of engagement projects. Ambassadors Adele Luta, Sydney Hamilton, and Samantha Porter spent 24 hours underwater to become aquanauts and hosted live-streamed chats about their experience. Neha Murad was one of nine ambassadors who helped create video episodes for GoldieBlox's online Curiosity Camp. Dorothy Tovar told the story of how she overcame her fear of nature as part of a Story Collider podcast organized by fellow ambassador Becca Peixotto. Other ambassadors developed lending libraries of STEM toys, picture books, and card games with their grants.

Health technologist Nicole Jackson used her grant to train 20 women scientists to add and edit Wikipedia entries for women scientists, generating 1.96 million page views so far. The trainees are now training other women contributors to exponentially expand the reach of the project, she said. "When girls wonder, 'Where are the women in STEM?,' we want them to see that they're already here and they're doing great work."

Since becoming an ambassador, Jackson has connected with 1,090 students from 26 states through the educational volunteering platform Nepris, sharing her own evolving experiences in health care product development during the pandemic with students.

Girls then reached out to her with questions about their own product development prototypes—along with more personal concerns, she said. "I remember the last one I did, we started talking about George Floyd. We started talking about the emotional impacts of being a Black woman leader in technology," she recalled.



2020 AAAS Annual Report
annualreport.aaas.org

"They wanted to hear the truth, and I think the fact that they were bold enough to ask tells you everything you need to know about this generation," Jackson added.

Gracie Ermi, a computer scientist who writes code to aid wildlife conservation, also used the Nepris program as an ambassador. One of her favorite experiences was speaking with students at her own elementary school. At that age, she was interested in science but "had no concept of what that kind of job looked like," she said, "so the goal of all the presentations I give is to be a face for this type of work."

Talking with students motivates her to work harder, she said. "A lot of them are driven to make a difference, especially in environmental areas, and the questions they ask are so thoughtful."

Ermi contributed her grant money toward a book project called Secret Lives of STEM Ambassadors, in which the women innovators answer questions from girls ages nine to 14. Girls sometimes wonder if they need to give up other interests or future goals to work in STEM, she said, "but one of our goals is to show that we are multifaceted people, that we are not just the jobs that we do."

"If we want kids to be more engaged...we should never introduce ourselves as our profession," Jackson agreed. "We should introduce ourselves as the people that we are and talk about the journey that took us through the things that interested us."

Opportunities for engagement "have exploded" since she became an ambassador, said Ermi, who has given more than 60 presentations, filmed an episode of GoldieBlox's Curiosity Camp, and was a guest on the Strong Voice podcast. Professional development workshops offered by AAAS, as well as sessions at the AAAS Annual Meeting, have been especially helpful in boosting her communication skills, she noted.

AAAS-organized trainings covered topics from intellectual property law to engaging with policy-makers and were based on the ambassadors' feedback, said Cloyd. She and her colleagues also have acted as sounding boards and coaches and have supported ambassadors at ComicCon and other science and entertainment conventions.

"We're bringing AAAS's expertise in science communication training but we're also able to shine a spotlight on ambassadors who have had these experiences, so that they can learn from one another," Cloyd said.

Ambassadors said the powerful peer-to-peer mentoring they get from each other is a crucial part of the program—Jackson called it "earth-shattering support." The ambassadors regularly consult each other on their own professional challenges, share engagement opportunities, and offer a rich network of expertise, participants said.

The program, originally scheduled to end in February 2021, has been extended to the end of the year to give the ambassadors more time to complete their grant projects. The ambassadors' profiles, photos, and videos are now part of the IF/THEN Collection (www.ifthencollection.org), one of the largest digital resources dedicated to increasing access to authentic images of women in STEM.



Partner with the global leader involved in over 50% of COVID-19 testing to help solve the most intractable crisis we face today.

We offer:

Accuracy and reliability—advanced assay designs compensate for current and emerging mutations like the Delta (B.1.617.2) and Alpha (B.1.1.7) variants

Service partnerships—help ensure minimal downtime, allowing operations to continue running at their peak

Single-source convenience—one supplier reliably providing all consumables, instruments, and servicing to enable uninterrupted testing

Keep moving forward in the fight against COVID-19 and learn about our offerings by scanning the QR code above.

Get a free infographic on lab-based and rapid PCR testing at **thermofisher.com/covid19scienceinfographic**

ThermoFisher SCIENTIFIC

© 2021 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. **COL25296 0721**

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



PHOTO: NASA IMAGE COURTESY JEFF SCHMALTZ/LANCE/EOSDIS MODIS RAPID RESPONSE TEAM AT NASA GSFC

PALEOCLIMATE

Cooling threshold

The end of each interglacial period over the past 800,000 years was characterized by large, abrupt cooling episodes that were distinct from the gradual decrease in insolation that was occurring. Why did these sudden coolings occur? Yin *et al*. used a suite of climate models to show that an insolation threshold exists, beneath which the slow decrease of energy from the Sun leads to rapid climate cooling. This effect results from weakening of the Atlantic meridional overturning circulation involving sea ice feedbacks in the Nordic and Labrador Seas. —HJS *Science*, abg1737, this issue p. 1035

Abrupt changes in global climate after interglacial periods are tied to gradual changes in insolation and feedbacks with sea ice in the Arctic.

QUALITY CONTROL

Safeguarding protein complex assembly

The assembly of multiprotein complexes inside the cell requires each subunit to be produced at a defined level relative to its partners. Imbalances in subunit synthesis are inevitable, necessitating the elimination of unassembled intermediates. Zavodszky *et al*. found that a ubiquitin ligase called HERC1 is responsible for marking certain assembly intermediates of the proteasome for degradation. HERC1 finds these intermediates by recognizing a proteasome assembly factor that normally dissociates when assembly is complete. A point mutation in HERC1 that impairs its ability to recognize proteasome assembly intermediates causes neurodegeneration in mice, highlighting the importance of this quality control pathway. —SMH

Science, abc6500, this issue p. 998

APPLIED PHYSICS

Thin, sensitive skin electronics

The properties of the human sense of touch, including high sensitivity to differences in temperature, pressure, or surface roughness, are challenging to replicate in robotics because skin materials must be highly conductive, stretchable, and thin. Jung *et al*. developed a process to assemble nanomaterials as a monolayer that is partially embedded in an ultrathin elastomer. The process works by depositing a mixed solvent containing nanostructured silver and/or gold, along with elastomer, onto deionized water. This results in a layer of nanoparticles residing at the interface coating with elastomer, which is further densified by the addition of surfactant. The process is scalable, and the resulting elastomer membranes can be transferred to other substrates. —MSL

Science, abh4357, this issue p. 1022

ORGANIC CHEMISTRY

Shuffling nitrogen with a light push

Manipulation of carbon–nitrogen rings is integral to the synthesis of numerous pharmaceutical and agrochemical compounds. Jurczyk *et al*. report that photoexcitation of carbonyl-substituted cyclic amines can shift the nitrogen from inside to outside the ring framework. The reaction appears to proceed through a 1,5-hydrogen shift to the electronically excited carbonyl, which sets in motion the subsequent carbon–nitrogen and carbon–carbon bonding rearrangements. Several oxygen and sulfur heterocycles were

applicable as well. Addition of a chiral phosphoric acid catalyst rendered the reaction asymmetric. —JSY

Science, abi7183, this issue p. 1004

PALEOECOLOGY

More species in warm waters

Past patterns of diversity can be used as a baseline for understanding how current human-induced changes are affecting biodiversity. Womack *et al*. looked at mollusk fossils from 40 million years ago in New Zealand to determine how ocean temperatures influence species richness and functional redundancy, a measure of how many species fill similar ecological roles. Both richness and redundancy increased in periods with warmer water, meaning that there were more species and that those species often filled similar ecological roles. Such ecological redundancy can increase ecosystem resilience, and understanding its relationship with temperature can help us determine where human activities are driving change. —SNV

Science, abf8732, this issue p. 1027

HUMAN GENETICS

Gauging the spectrum of human mutations

It has become increasing clear that mutation affects phenotypic variation and disease risk across humans. However, there are many different types of mutation. Seplyarskiy *et al*. applied a matrix factorization method to large human genomic datasets to identify germline mutational processes in an unsupervised manner. From this survey, nine robust mutational components were identified and specific mechanisms generating seven of these processes were proposed from correlations with genomic features. These results confirm and improve upon our understanding of mutational processes and reveal likely mechanisms of mutation in the human genome. —LMZ

Science, aba7408, this issue p. 1030

IMMUNOLOGY

Division of labor for vaccine responses

Modified vaccinia Ankara (MVA) is an attenuated pox virus vaccine that may be a safe vector for vaccines against other viruses. Doring *et al*. analyzed how human dendritic cells (DCs), which are critical for vaccine efficacy, responded to MVA. They found a division of labor among DCs in which infected DCs produced inflammatory cytokines that activated nearby noninfected DCs, which in turn expressed T cell–costimulatory molecules. The subsequent and rapid death of infected DCs occurred in a parallel but independent manner from that of the cytokine response. —LKF

Sci. Signal. **14**, eabd9720 (2021).

ARCHAEOLOGY

Insights into the Roman diet from Herculaneum

Both historical sources and standard stable isotope analyses of human remains have offered insights into patterns of food consumption in ancient Rome. However, this research is often not of a sufficiently high resolution to explore variations in diet between the sexes, in different occupations, or with more subtle degrees of social status. To address this shortcoming, Soncini *et al.* used a combination of stable isotope analysis of bone collagen with Bayesian models of protein synthesis of 17 adults who perished in the 79 CE eruption of Mount Vesuvius. They found that men consumed more fish than women, but in terms of total energy obtained, the diets of both sexes were dominated by cereals, terrestrial animals, olives, and possibly wine. —MSA

Sci. Adv. 10.1126/sciadv.abg5791 (2021).

IN OTHER JOURNALS

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



IMAGE: DENNIS KUNKEL MICROSCOPY/SCIENCE SOURCE

NEUROSCIENCE

Selective pruning of synapses

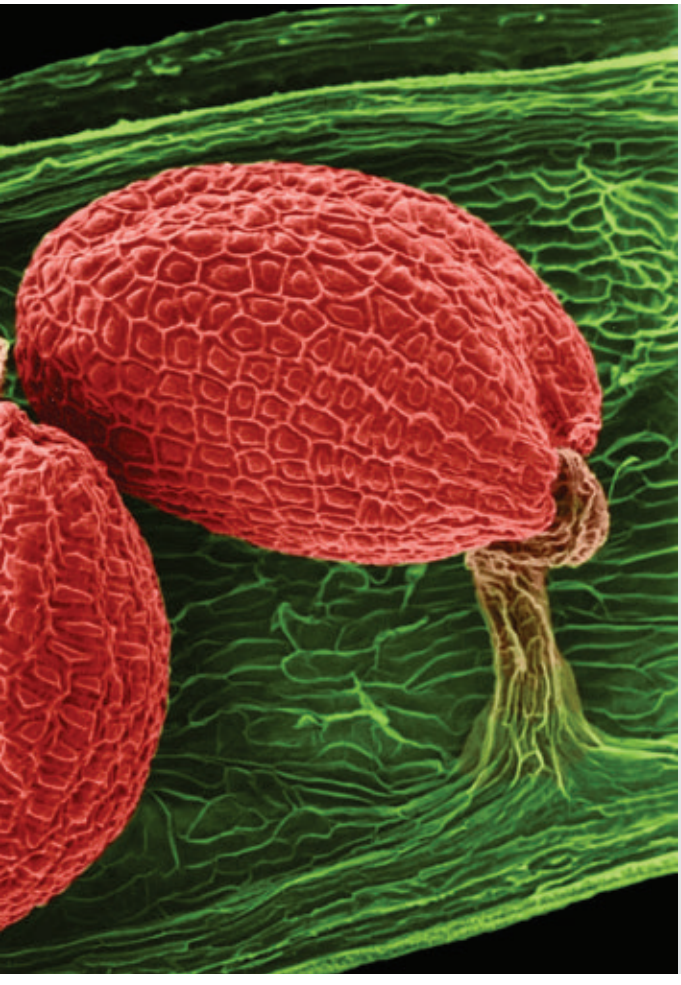
A fundamental question in developmental neuroscience is how different cell types wire together with exquisite specificity to ensure the formation of canonical neural circuits. Increasingly, non-neural cells have been implicated as being essential to this process. Microglia-resident immune cells of the brain play crucial roles in refining synaptic connections. Favuzzi *et al.* show that GABA-receptive microglia interact with inhibitory synapses in developing mice 2 to 3 weeks after birth. Within this population of microglia, GABA promotes the selective pruning of inhibitory connectivity. Perturbing these specialized microglia caused long-lasting defects in inhibitory connectivity without affecting excitatory synapses and led to hyperactivity in adult animals. Thus, distinct microglial populations differentially engage with specific synapse types during development to modulate behavior. —SMH

Cell **184**, P4048 (2021).

PHOTOACTIVATED BONDS

Visibly controlling bond breaking

Photodynamic bonds that are stable under dark conditions but break when photoexcited enable an array of materials applications where switching is desired. For cycloadditions and reversions of conjugated organic molecules or radical formation, the excitation energy to activate bonds usually requires ultraviolet light. Han *et al.* report that the coordination of a selenoether ligand



PLANT SCIENCE

Separating the stages of germination

Many plant seeds exist in a desiccated state awaiting water for germination. Dorone *et al.* identified the intrinsically disordered protein FLOE1 in the *Arabidopsis* seed proteome. FLOE1 remains distributed in a desiccated environment but forms condensates upon hydration in a reversible process that is sensitive to changes in water potential. Natural variation in the aspartate-serine–rich domain fine-tunes the response to hydration. In vivo, this phase separation of FLOE1 coincides with seed germination. Genetic and allelic diversity seems to position FLOE1 in bet hedging for an individual plant, as well as in adaptation across plant lineages to varying niches, all to optimize the timing of seed germination. —PJH

Cell **184**, P4284 (2021).

Rehydration triggers protein condensation within *Arabidopsis* seeds.

with a ruthenium coordination complex can be reversed by breaking the ruthenium–selenium bond with visible light and can also be restored after dark storage. They applied this approach to photocontrol of amphiphiles between spherical micelles and bowl shapes, wetting of surfaces, and polymer gelling. —PDS

J. Am. Chem. Soc. 10.1021/jacs.1c05648 (2021).

ORGANIC CHEMISTRY

A trio of pyrroles

Macrocycles encompassing four pyrrole rings are common motifs in biochemical and synthetic light harvesting and catalysis. Inaba *et al.* report a strained, smaller analog with three pyrroles. Their route cyclized a linear hexaketone precursor through condensation, restored the six carbonyls through an oxidation–reduction sequence, and then aminated it to produce the pyrroles. Crystallography highlighted the high strain inherent in the structure, which quickly underwent a ring-opening dimerization in acidic solution to form a calix[6]pyrrole. The results help to explain why the tetrapyrrole structure is the smallest of these macrocycles that is readily observed. —JSY

J. Am. Chem. Soc. 10.1021/jacs.1c06331 (2021).

FETAL BIOLOGY

Growing importance of genetics

An infant's birth weight is associated with a variety of health outcomes all the way into adulthood. Birth weight is influenced by both fetal and maternal factors, as well as by environmental and genetic regulation. To elucidate the relative contributions of genetic factors from mother and child, Juliusdottir *et al*. performed genome-wide association studies on hundreds of thousands of mothers and newborns from the Icelandic birth register and other databases. Most of the genes that affected growth were part of the fetal genome, but the authors also identified some parental alleles that had specific effects, as well as pathways influenced by these genes, such as glycemic control and cardiovascular function. —YN

Nat. Genet. **53**, 1135 (2021).

CATALYST DESIGN

pH-universal OER electrocatalyst

The oxygen evolution reaction (OER) is the key electrocatalytic process in many electrochemical systems, including water splitting and metal–air batteries. Because of the acid-aggressive and strongly corrosive media, the development of electrocatalysts for OER applications that could efficiently operate over the long term is an outstanding problem. Zhang *et al.* developed a sodium-doped amorphous, crystalline RuO_2 with rich oxygen vacancies that displays high catalytic stability and competitive OER activity over a wide pH range, including strongly acidic conditions. Density functional theory calculations revealed that the underlying mechanism is possible in other similar systems, implying that this work offers a new pathway with which to develop efficient OER electrocatalysts under pH-universal conditions. —YS

Angew. Chem. Int. Ed. Engl. **60**, 18821 (2021).

MOLECULAR BIOLOGY

Mirror image DNA

In a world barely nascent, scientists can create mirror images of the building blocks of life, including DNA, RNA, and protein, and use them to build a mirror image of life itself. This fantastic endeavor is not an easy task. A major challenge is to create versions of enzymes, the workhorses for biosynthesis in cells, as mirror images. Fan *et al.* created a high-fidelity mirror image *Pfu* DNA polymerase. They used this enzyme to assemble a stretch of mirror image DNA containing 1500 mirror image nucleotides. The idea is that this DNA can be used as secure information storage to evade the threat of degradation by enzymes in natural organisms and the environment. Mirror image DNA survives much longer in environmental water samples than its natural counterpart. The next step is to find out how mirror biotechnology can be used to bring real-life benefits. —DJ

Nat. Biotechnol. 10.1038/s41587-021-00969-6 (2021).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

The case for wildlife as the COVID-19 origin

Despite numerous hypotheses for how COVID-19 emerged in humans in Wuhan, China, in the winter of 2019, many experts have surmised that this was a zoonotic disease that originated in horseshoe bats. In a Perspective, Lytras *et al.* discuss the evidence supporting COVID-19 as a spillover event, likely from an intermediate host species that acted as a conduit from horseshoe bats, which are a host reservoir for severe acute respiratory syndrome-related coronaviruses (SARSr-CoVs). The authors discuss the events at the time that increased the likelihood of such a spillover, including an increased demand for meat in the wake of a pig farm culling associated with an outbreak of the African swine fever virus. Rising demand increased farming of, contact with, and encroachment into the habitats of wildlife that are susceptible to SARSr-CoVs. The SARS epidemic in 2002 revealed the danger posed by SARSr-CoVs, and the current SARS-CoV-2 pandemic emphasizes again that these viruses pose an immediate risk that must be monitored. —GKA

Science, abh0117, this issue p. 968

STRUCTURAL BIOLOGY

Bright future ahead for crystallography

Macromolecular x-ray crystallography typically provides static snapshots of systems at equilibrium. Advances in time-resolved crystallography have made it possible to capture dynamics in biomolecules: large and small, fast and slow. Brändén and Neutze review techniques and concepts that have emerged from recent work at x-ray free electron laser sources and are now being applied in other settings and to a growing number of biological systems. Despite challenges in analyzing and relating these data to a biological context, experiments in this field have opened new frontiers in temporal and spatial resolution and yielded many new insights into nonequilibrium chemistry and conformational changes in biology. —MAF

Science, aba0954, this issue p. 980

CORONAVIRUS

Mechanisms of airborne transmission

The COVID-19 pandemic has highlighted controversies and unknowns about how respiratory pathogens spread between hosts. Traditionally, it was thought that respiratory pathogens spread between people through large droplets produced in coughs and through contact with contaminated surfaces (fomites). However, several respiratory pathogens are known to spread through small respiratory aerosols, which can float and travel in air flows, infecting people who inhale them at short and long distances from the infected person. Wang *et al*. review recent advances in understanding airborne transmission gained from studying the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections and other respiratory pathogens. The authors suggest that airborne transmission may be the dominant form of transmission for several respiratory pathogens, including SARS-CoV-2, and that further understanding of the mechanisms underlying infection from the airborne route will better inform mitigation measures. —GKA

Science, abd9149, this issue p. 981

IMMUNOLOGY

Mom's IL-6 rewires baby's gut immunity

Most infections that occur during pregnancy are mild and transient. However, whether such pathogen encounters can shape the long-term trajectory of the offspring's immune system remains unclear. Lim *et al*. infected pregnant mice with the common food-borne pathogen *Yersinia pseudotuberculosis* (YopM) (see the Perspective by Amir and Zeng). Although the infection was maternally restricted and short-lived, the offspring harbored greater numbers of intestinal T helper 17 cells into adulthood. Interleukin-6 (IL-6) mediated this tissue-restricted effect by acting on the fetal intestinal epithelium during development. Although offspring from mothers infected with YopM or injected with IL-6 showed enhanced resistance to oral infection with *Salmonella* Typhimurium, they also exhibited higher susceptibility toward enteric inflammatory disease. —STS

Science, abf3002, this issue p. 982; see also abl3631, p. 967

NEURODEVELOPMENT

The cerebellum reveals its genetic programs

Gene-regulatory networks govern the development of organs. Sarropoulos *et al*. analyzed mouse cerebellar development in the context of gene-regulatory networks. Single nuclear profiles analyzing chromatin accessibility in about 90,000 cells revealed diversity in progenitor cells and genetic programs guiding cellular differentiation. The footsteps of evolution were apparent in varying constraints on different cell types. —PJH

Science, abg4696, this issue p. 983

CORONAVIRUS

A broad defense against SARS-like viruses

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the third coronavirus that has emerged as a serious human pathogen in the past 20 years. Treatment strategies that are broadly protective against current and future SARS-like coronaviruses are needed. Martinez *et al*. took on this challenge by developing vaccines based on chimeras of the viral spike protein. The messenger RNA vaccines encode spike proteins composed of domain modules from epidemic and pandemic coronaviruses, as well as bat coronaviruses with the potential to cross to humans. In aged mice vulnerable to infection, the chimeric vaccines protected against challenge from SARS-CoV, SARS-CoV-2 and tested variants of concern, and zoonotic coronaviruses with pandemic potential. —VV

Science, abi4506, this issue p. 991

MOLECULAR BIOLOGY

RNA editing restricts ciliary kinases

Ciliary kinases are essential for cilia formation and function but it remains unknown how their activities are regulated in vivo. Li *et al*. created roundworm animal models carrying hyperactive ciliary kinases that disrupt cilia. Their genetic suppressor screens revealed that loss of an RNA adenosine deaminase that catalyzes adenosine-to-inosine (A-to-I) RNA editing rescued ciliary abnormalities. They found that kinase hyperactivation caused this RNA adenosine deaminase to edit kinase RNA and impair kinase RNA splicing and translation, thereby down-regulating ciliary kinases from nuclei. These results suggest that ciliopathies may be treated by targeting the pathways outside of cilia. —DJ

Science, abd8971, this issue p. 984

SUPERCONDUCTIVITY

Not your usual superconductor

Most superconductors have only one superconducting

phase. Khim *et al*. measured the magnetic susceptibility of the heavy fermion material $CeRh_2As_2$ to reveal the presence of two distinct superconducting phases, one of which emerges from the other when an external magnetic field is applied (see the Perspective by Pourret and Knebel). The researchers ascribe the unusual properties of $CeRh_2As_2$ to its crystal structure, which is globally centrosymmetric but consists of noncentrosymmetric layers. —JS

Science, abe7518, this issue p. 1012; see also abj8193, p. 962

THERMAL TRANSPORT

Blocking heat in two ways

Low thermal conductivity is important for barrier coatings, thermoelectrics, and other applications. Gibson *et al*. combined two complementary methods that manipulate internal interface properties to dramatically decrease the thermal conductivity of the inorganic material BiO_2Cl_2Se (see the Perspective by Kim and Cahill). The authors took advantage of both in-plane structural distortions and weak bonding layers to push the conductivity down to 0.1 watts per kelvin per meter, which is only four times that of air. The principles should be applicable to other systems and provide a method for developing crystals with extremely low thermal conductivity. —BG

Science, abh1619, this issue p. 1017; see also abk1176, p. 963

MICROBIOME

SagA promotes immunotherapy response

The gut microbiome can influence the treatment outcome for cancer patients receiving PD-L1 immunotherapy, but the mechanisms underlying favorable responses are unclear. Griffin *et al*. found that a particular type of bacteria called enterococci enhance anti–PD-L1 immunotherapy in mice (see the Perspective by Ansaldo and Belkaid). The researchers show that enterococci secrete an enzyme called SagA that breaks down components of the bacterial cell wall. This process results in the release of muramyl peptide fragments, which in turn act as stimulatory molecules to promote signaling of the innate immune sensor protein NOD2 and improved immunotherapy responses. —PNK

Science, abc9113, this issue p. 1040; see also abl3656, p. 966

RNA

Machine learning solves RNA puzzles

RNA molecules fold into complex three-dimensional shapes that are difficult to determine experimentally or predict computationally. Understanding these structures may aid in the discovery of drugs for currently untreatable diseases. Townshend *et al*. introduced a machine-learning method that significantly improves prediction of RNA structures (see the Perspective by Weeks). Most other recent advances in deep learning have required a tremendous amount of data for training. The fact that this method succeeds given very little training data suggests that related methods could address unsolved problems in many fields where data are scarce. —DJ

Science, abe5650, this issue p. 1047; see also abk1971, p. 964

CANCER

Taking a BiTE out of tumors

Bispecific T cell engagers (BiTEs) and other forms of T cell–engager therapies represent a promising, off-the-shelf treatment for many types of cancer. However, it is unclear why some types of tumors are particularly resistant to BiTE treatment. Belmontes *et al.* evaluated pharmacokinetics, pharmacodynamics, and immune correlates of responsiveness to BiTE treatment. In multiple mouse tumor models, they found that pretreatment T cell density in tumors was associated with BiTE efficacy. In T cell–cold tumors, which lack abundant T cells, coadministration of immune checkpoint blockade antibodies or a 4-1BB agonist overcame resistance to BiTE treatment. Together, these findings identify combination therapeutic strategies for BiTES and other T cell engagers. —CSM

Sci. Transl. Med. **13**, eabd1524 (2021).

IMMUNOTHERAPY

TCR-like antibodies tackle celiac disease

Ingestion of gluten-containing food triggers the gastrointestinal symptoms of celiac disease in patients with $CD4^+$ T cells specific for deamidated gluten peptides presented by cells. Frick *et al.* used phage display technology to screen for T cell receptor (TCR)–like antibodies specific for an immunodominant gluten peptide bound by the protein HLA-DQ2.5. Antibody engineering optimized affinity and binding stability, yielding an improved TCR-like antibody that structurally mimicked the TCR interface with gluten peptide–protein complexes. These TCR-like antibodies blocked the activation and proliferation of gluten-responsive human $CD4^+$ T cells both in vitro and in DQ2.5-transgenic mice. TCR-like antibodies that block immunodominant epitope recognition have potential as personalized treatments for blunting gluten-activated T cell responses without compromising the effector functions provided by other T cells. —IRW

Sci. Immunol. **6**, eabg4925 (2021).

REVIEW SUMMARY

STRUCTURAL BIOLOGY

Advances and challenges in time-resolved macromolecular crystallography

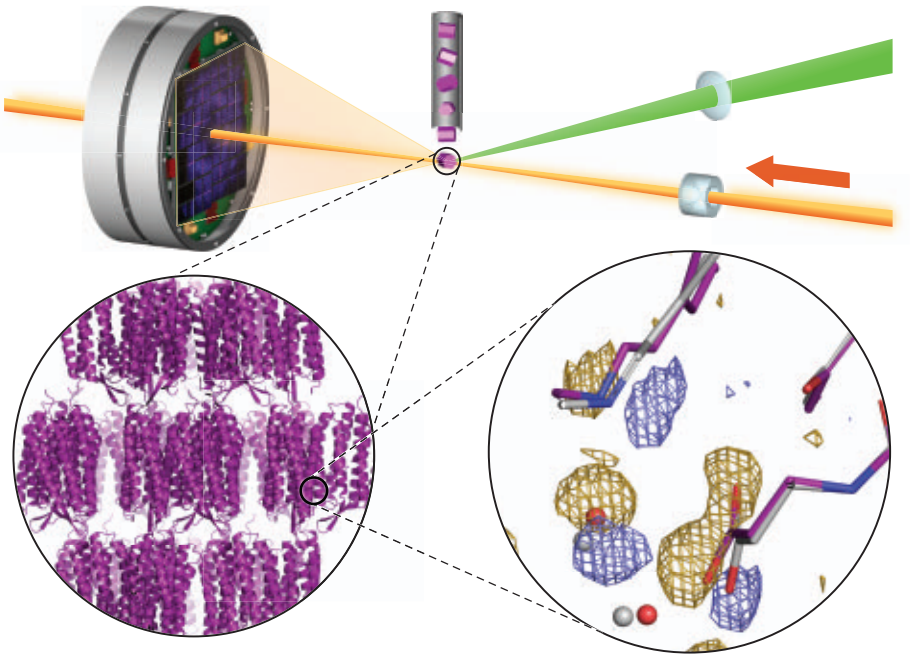
Gisela Brändén and Richard Neutze*

BACKGROUND: Conformational changes are essential for the correct functioning of biological macromolecules. Time-resolved x-ray crystallography extends an extremely successful method for the structural determination of biomolecules by incorporating time as a fourth dimension. Time-resolved x-ray diffraction studies are performed at room temperature so as to allow the biological reaction to evolve within crystals. This reaction must also be initiated throughout crystals—and x-ray diffraction data must be collected—at least as rapidly as the fastest time point of interest. Mature structural analysis tools of macromolecular crystallography can then be adapted to allow x-ray diffraction data to be analyzed in terms of time-dependent conformational changes.

ADVANCES: Decades of work using polychromatic x-ray pulses (Laue diffraction) at synchrotron radiation sources laid the experimental foundations underpinning the field of time-resolved macromolecular crystallography. Light-driven biological reactions can be rapidly initiated throughout crystals using short laser pulses and have therefore been a major focus for the field. Serial crystallography was first demonstrated 10 years ago at an x-ray free-electron laser (XFEL). In this approach, x-ray diffraction data are collected from a sequence of microcrystals, typically 10 μm or less in their largest dimension, that are being continuously replaced. X-ray diffraction data from thousands of microcrystals are then merged into a complete dataset. Sample delivery for serial crystallography experiments at an XFEL initially relied on liquid microjets, but many other sample-delivery technologies have since been developed, each with its own strengths and weaknesses. Serial crystallography has overcome many of the technical limitations that constrained time-resolved Laue diffraction and has thereby transformed the field, creating a renaissance of interest in time-resolved macromolecular crystallography.

In this Review, we describe how time-resolved x-ray diffraction studies using XFEL pulses a few femtoseconds (10^{-15} s) in duration have allowed atomic motions in biological samples to be visualized on the time scales at which chemical bonds break or isomerize or at which electrons move. We illustrate the power of time-resolved serial crystallography to yield structural and functional insights on slower time scales by showcasing structural results from two energy-transducing membrane proteins, bacteriorhodopsin and photosystem II, neither of which were amenable to synchrotron-based time-resolved Laue diffraction. We also discuss structural results obtained when using mixing jets to diffuse reactants into microcrystals or when releasing photocaged compounds using a laser flash, which have allowed biological reactions that are not naturally light sensitive to be followed in time.

OUTLOOK: Time-resolved crystallography is transitioning from a highly technical domain of specialists into a flexible approach for elucidating structural and functional insights from macromolecules in their crystalline state. Although serial crystallography was first developed for XFEL-based studies, the recent transfer of time-resolved serial crystallography to synchrotron radiation facilities is critical for the growth of the user community. Nonspecialist user communities will also drive standardized experimental setups that further lower entry barriers for new users. Structural conclusions drawn from XFEL-based studies of ultrafast structural changes should be consolidated by repeating these experiments using lower power density photoexcitation conditions, and data analysis steps—from processing experimental data through to structural interpretation—need to be streamlined. As further structural insights emerge from an increasingly diverse set of macromolecules, it becomes possible to imagine a time when structural results from time-resolved diffraction experiments become as central to understanding a biological reaction as the resting-state structure of a macromolecule is today. ■



Schematic illustration of a time-resolved serial crystallography study of a light-sensitive protein. A continuous stream of microcrystals (purple) is injected across a focused x-ray beam (orange) at a synchrotron or XFEL facility. Diffraction data collected from microcrystals photoactivated by a laser pulse (green) are compared against reference diffraction data without photoactivation. Electron density changes (inset: positive difference electron density, blue; negative, yellow) are modeled as changes in protein structure for each time delay between laser and x-ray pulse, providing structural insight into the biological reaction.

Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden.
*Corresponding author. Email: richard.neutze@gu.se
Cite this article as G. Brändén, R. Neutze, *Science* **373**, eaba0954 (2021). DOI: 10.1126/science.aba0954

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aba0954

REVIEW

STRUCTURAL BIOLOGY

Advances and challenges in time-resolved macromolecular crystallography

Gisela Brändén and Richard Neutze*

Conformational changes within biological macromolecules control a vast array of chemical reactions in living cells. Time-resolved crystallography can reveal time-dependent structural changes that occur within protein crystals, yielding chemical insights in unparalleled detail. Serial crystallography approaches developed at x-ray free-electron lasers are now routinely used for time-resolved diffraction studies of macromolecules. These techniques are increasingly being applied at synchrotron radiation sources and to a growing diversity of macromolecules. Here, we review recent progress in the field, including visualizing ultrafast structural changes that guide the initial trajectories of light-driven reactions as well as capturing biologically important conformational changes on slower time scales, for which bacteriorhodopsin and photosystem II are presented as illustrative case studies.

Time-resolved macromolecular x-ray crystallography allows for visualization of the movements of atoms over time and thereby provides insight into the chemical pathways underpinning biological reactions. A successful time-resolved diffraction experiment requires well-diffracting crystals of the target enzyme, a means to initiate a reaction throughout these crystals, a method of rapidly collecting diffraction data, and tools for analyzing x-ray diffraction data in terms of time-dependent conformational changes. The extent to which the reaction is synchronized and the duration of the x-ray exposures set a lower bound on the reaction time scale that can be probed. Electron density changes due to an enzyme-catalyzed reaction occurring within crystals were first visualized several decades ago (*1*–*3*), yet numerous technical challenges have hindered the widespread application of time-resolved crystallography (*4*). Serial crystallography (*5*), in which x-ray diffraction data are collected from a sequence of microcrystals that are being continuously replaced, has developed in tandem with x-ray free-electron laser (XFEL) sources (*6*) and has sparked renewed excitement in time-resolved diffraction as it has proved relatively straightforward to adapt serial data-collection strategies to the study of protein structural changes (*7*).

To begin, we briefly outline historical developments underpinning both time-resolved diffraction and serial crystallography. We then discuss how XFEL pulses a few femtoseconds in duration have created new possibilities for studying ultrafast protein dynamics. We illustrate the power of time-resolved serial crystallography on slower time scales by highlighting structural results from two light-driven, energy-transducing membrane proteins, bacteriorhodopsin and photosystem II, neither of which are amenable to more traditional time-resolved diffraction approaches. We also outline recent successes using caged compounds or mixing schemes to initiate enzymatic reactions within microcrystals, which extend the sphere of application beyond naturally light-driven systems. These advances bode well for time-resolved structural enzymology to increasingly be used to characterize the functional mechanisms of biological macromolecules.

Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden.
*Corresponding author. Email: richard.neutze@gu.se

Time-resolved Laue diffraction and time-resolved x-ray solution scattering

When determining a static x-ray structure of a macromolecule, diffraction data are typically collected using a monochromatic x-ray beam from a crystal that is cooled to cryogenic temperature. For time-resolved studies, it is essential that the reaction proceeds within a crystal, which means that diffraction data are collected at room temperature. Moreover, each diffraction frame must be collected more rapidly than the reaction of interest. This requirement motivated a consideration of Laue diffraction approaches, whereby a crystal is exposed to a polychromatic x-ray beam (*8*, *9*) and consequently more diffraction spots are sampled. Laue diffraction allows data to be collected using much shorter x-ray exposures (*4*), but the method is extremely sensitive to crystalline disorder.

Pioneering time-resolved Laue diffraction experiments used the photolysis of caged guanosine triphosphate (GTP) to record time-dependent electron density changes from crystals of Ha-Ras p21 (*3*) and achieved a time resolution of minutes. A sequence of time-resolved Laue diffraction studies using crystals of myoglobin with carbon monoxide bound to the active site heme (Mb:CO) (*10*, *11*) and photoactive yellow protein (PYP) (*12*–*15*) later extended this time resolution into the millisecond (*12*), nanosecond (*10*, *13*), and picosecond (*11*, *14*, *15*) regimes using increasingly powerful synchrotron radiation beams and higher performance x-ray detectors. Conformational changes in Mb:CO and PYP are initiated using a laser flash, both proteins return to their resting conformation within seconds, and both proteins yield very well-diffracting crystals that are resistant to disorder. Very few proteins fulfill all of these requirements, and, consequently, time-resolved Laue diffraction successes have been reported for only a handful of other proteins, including dimeric hemoglobin (*16*), tetrameric hemoglobin (*17*), the heme domain of the oxygen sensor FixL (*18*), and a photosynthetic reaction center (*19*, *20*). One highly creative recent study used very strong electric fields (~100 kV/mm) in combination with time-resolved Laue diffraction to visualize electric field–induced conformational changes in crystals of a PDZ domain of human E3 ubiquitin ligase LNX2 (*21*).

As the fruit of near-Herculean efforts spanning several decades that addressed technical challenges (*4*), time-resolved Laue diffraction established that electron density changes in proteins can be visualized in real time and that the trajectories of dissociated ligands, the isomerization of a buried cofactor, and the response of surrounding side chains can be measured experimentally. An intrinsic limitation of time-resolved x-ray diffraction is that motions that conflict with the crystal packing cannot be visualized and, in some cases, the natural course of a reaction may be slowed or suppressed. To some extent, this limitation has been addressed by developing time-resolved x-ray solution scattering (TR-XSS) (*22*–*26*), which measures secondary structural rearrangements from proteins in solution. TR-XSS is not as information rich as x-ray diffraction and, consequently, structural modeling is performed as low-resolution shape reconstructions (*25*, *26*) or relies on additional information such as deposited crystallographic structures (*22*, *23*) or the output from molecular dynamics simulations (*26*).

Serial femtosecond crystallography

Five decades of collaboration between structural biologists and synchrotron radiation facilities (*27*) highlight the interdependence of x-ray sources and their users. A disruptive technology emerged 11 years ago when the Linac Coherent Light Source (LCLS) was the first XFEL to lase at a wavelength of 1.5 Å (*6*). An XFEL is a revolutionary source because it allows extremely short and intense x-ray pulses to be focused upon very small samples (Box 1 and Fig. 1) and thereby facilitates major advances across many fields of science. Simulations predicted that the atoms of a biological sample would ionize during the course of an XFEL

exposure and the sample would be destroyed, yet interpretable diffraction data could be recovered if the XFEL pulse had passed through the sample before it had time to explode (*28*). This idea, later coined as "diffraction before destruction" (*29*), was verified experimentally by observing how the diffraction power falls off as protein crystals are exposed to XFEL pulses from 70 to 300 fs (1 fs = 10^{-15} s) in duration (*30*).

For macromolecular crystallography studies, the problem of sample destruction was addressed by continuously injecting a stream of micrometer-sized crystals (microcrystals) across a focused XFEL beam using a microjet (*31*, *32*) (Box 2). Diffraction data were read out at the repetition rate of the XFEL source (*5*), and data from a large number of microcrystals were merged together to recover complete diffraction data (*33*). This approach was called serial femtosecond crystallography (SFX) (*34*), because data were collected in a serial fashion using XFEL pulses a few femtoseconds in duration. The first SFX structures were determined at low resolution (*5*, *35*), but the method was extended to high resolution (*36*–*38*) as soon as an experimental station operating with an x-ray wavelength near 1 Å was commissioned (*39*). Several SFX structures have since been solved to high resolution (*40*–*42*), crystallographic phasing with XFEL radiation has been demonstrated (*43*–*46*), and phasing approaches that exploit the information recorded between diffraction spots have been described (*47*, *48*).

Time-resolved serial femtosecond crystallography

Serial crystallography addresses many of the technical hurdles that limited the broader application of time-resolved Laue diffraction. First and foremost, because microcrystals are continuously replaced (Box 2 and Fig. 2A), serial crystallography completely removes the constraint that a system of study must return to its resting state in order to cycle data-collection protocols. The extreme sensitivity of Laue diffraction to crystal disorder was also sidestepped by using quasi-monochromatic XFEL pulses, which typically have a photon energy bandwidth ~0.1% of the x-ray photon energy, almost two orders of magnitude narrower than the range often used in Laue diffraction studies at synchrotron sources. Conversely, new challenges arose in working with these data (*49*) owing to large shot-to-shot energy and spectral fluctuations in the XFEL source (*6*), variations in microcrystal quality, the random sampling of crystal orientation, and the fact that crystals do not have time to rotate during a femtosecond exposure and thus only partial diffraction intensities are collected. A landmark demonstration of time-resolved SFX (TR-SFX) used microcrystals of PYP to

Box 1. Synchrotron and x-ray free-electron laser radiation.

When electrons oscillate, they radiate electromagnetic radiation. This principle is used in all broadcasting devices, including television antennae and mobile telephones. At large x-ray user facilities, an electron bunch is accelerated to relativistic energies and passed through a periodic array of magnets (undulators) that are aligned in opposite directions within an array (Fig. 1). These alternating magnetic fields cause passing electrons to oscillate, and their relativistic energies mean that x-rays are generated. Synchrotrons are circular machines typically 500 to 1400 m in circumference, and most experimental instruments use undulators that are <5 m long. XFELs are driven by linear accelerators from 700 to 3400 m long that direct electrons into undulators typically 50 to 150 m in length. Linear accelerators produce far more compact electron bunches (~10 fs in duration) than do synchrotrons (~100 ps in duration), and hence XFEL pulses are several orders of magnitude shorter. The longer undulators used at XFELs facilitate a feedback mechanism whereby x-rays produced early in the undulator interact with the electron bunch, and a coherent process of lasing arises. Lasing produces many orders of magnitude more x-ray photons, with XFELs operating at ~1-Å wavelength usually delivering 10^{11} to 10^{12} x-ray photons per pulse for user experiments.

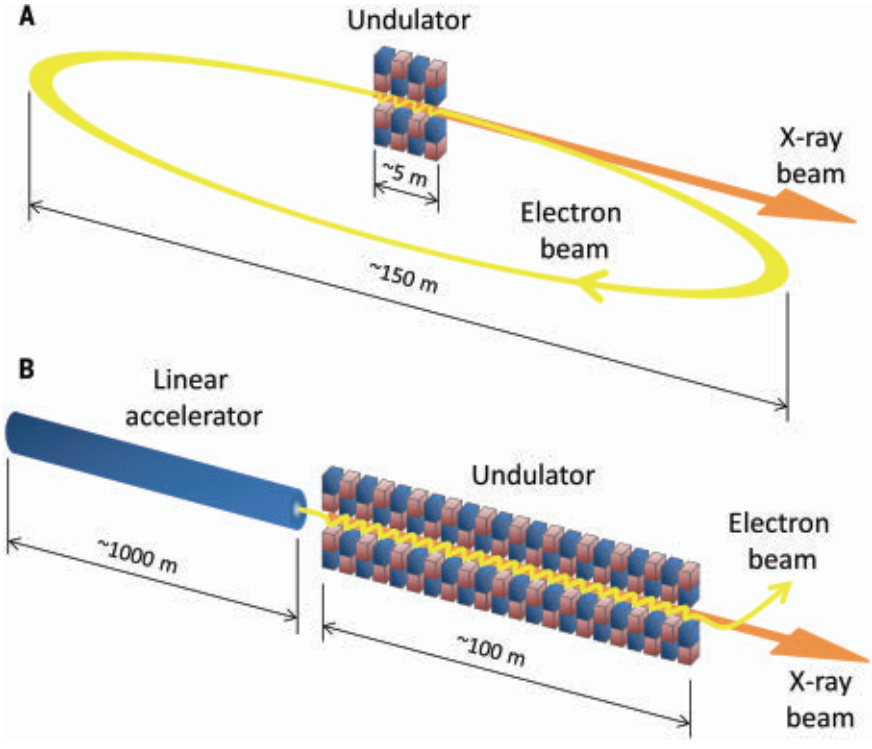


Fig. 1. Schematic illustrating how x-rays are generated at user facilities. (**A**) A synchrotron radiation source. (**B**) An XFEL source. Approximate length scales are indicated.

Box 2. Sample delivery.

In conventional macromolecular crystallography, a single crystal is cooled to cryogenic temperature, exposed to a monochromatic x-ray beam, and rotated to sample diffraction from multiple orientations. A single exposure to a focused XFEL beam will destroy a protein crystal (*5*, *28*), and therefore the sample must be continuously replaced. One approach is a liquid microjet (*31*, *32*) that continuously injects a stream of microcrystals across an XFEL beam (*5*). Innovative sample-delivery developments have realized microjet injectors for viscous samples (*118*) and increasingly sophisticated mix-and-inject systems (*97*, *98*, *101*, *103*, *119*–*122*) that blend a solution containing microcrystals with another containing substrate or other reagents before delivery across an x-ray beam. Fixed-target platforms are increasingly popular and involve premounting microcrystals on a chip that is physically translated between each x-ray exposure (*81*, *100*, *110*–*112*, *123*, *124*). Continuously circulating tape has also been used to pass microcrystal droplets through an x-ray beam (*87*, *99*). As these platforms mature, it seems likely that user convenience will lead to widespread use of a smaller number of standardized sample-delivery systems.

Box 3. Structural interpretation.

In time-resolved diffraction studies, a crystal in which a reaction has been initiated behaves as an ordered ensemble of both protein molecules at rest and a subpopulation that evolves with time. Given that the latter subpopulation is only a fraction of the total (*f*, called the activated-state occupancy), comparisons are normally made between diffraction data collected from crystals in which the reaction has been initiated (activated data) and data collected from crystals without the reaction being initiated (reference data). When both datasets have close to identical unit cell parameters, an isomorphous F_{obs}(activated) − F_{obs}(reference) difference Fourier electron density map is calculated by subtracting the scaled structure factor amplitudes of the reference dataset, F_{obs}(reference), from the scaled structure factor amplitudes of the activated dataset, F_{obs}(activated), using a structural model refined against the reference data for phases. Difference Fourier electron density maps (e.g., Fig. 2, C to H) are very sensitive to small changes in electron density (*50*) and are not biased by the phases of the active conformation. Positive difference electron density (blue) features indicate where new electron density arises for the activated state, whereas negative difference electron density (yellow) indicates where electron density from the resting state is depleted. The diffraction quality of the crystals and the activated-state occupancy are key experimental parameters to optimize, and very high multiplicity data should be collected in serial crystallography studies to average out noise contributions inherent to individual measurements (*71*). Bayesian schemes for estimating structure factor amplitude differences (*114*) and singular value decomposition protocols (*83*, *115*) have also improved the signal-to-noise ratio in difference Fourier electron density maps.

Whereas complementary positive and negative difference electron density features indicate the regions of a protein that move, the magnitudes of the atomic displacements must be modeled. This step often uses partial-occupancy structural refinement in which the resting-state conformation is held fixed as the coordinates and B-factors of the active-state conformation are refined with occupancy *f* against F_{obs}(activated) (*70*). Alternatively, the active-state conformation may be refined against data extrapolated to approximate that recorded from a hypothetical crystal with full occupancy, whereby F_{obs}(extrapolated) = $1/f \times F_{obs}$(activated) + $(1 - 1/f) \times F_{obs}$(reference) (*125*). Both structural refinement protocols involve compromise because partial occupancy refinement depends heavily on the quality of the resting-state model, whereas linear extrapolation of x-ray diffraction data exaggerates experimental errors (*125*). Coordinate errors can be estimated from an ensemble of activated-state conformations refined against a set of crystallographic observations with simulated noise (*82*) or by refining against a set of serial crystallography observations generated by resampling a subset of the measured diffraction intensities (*83*, *88*, *126*).

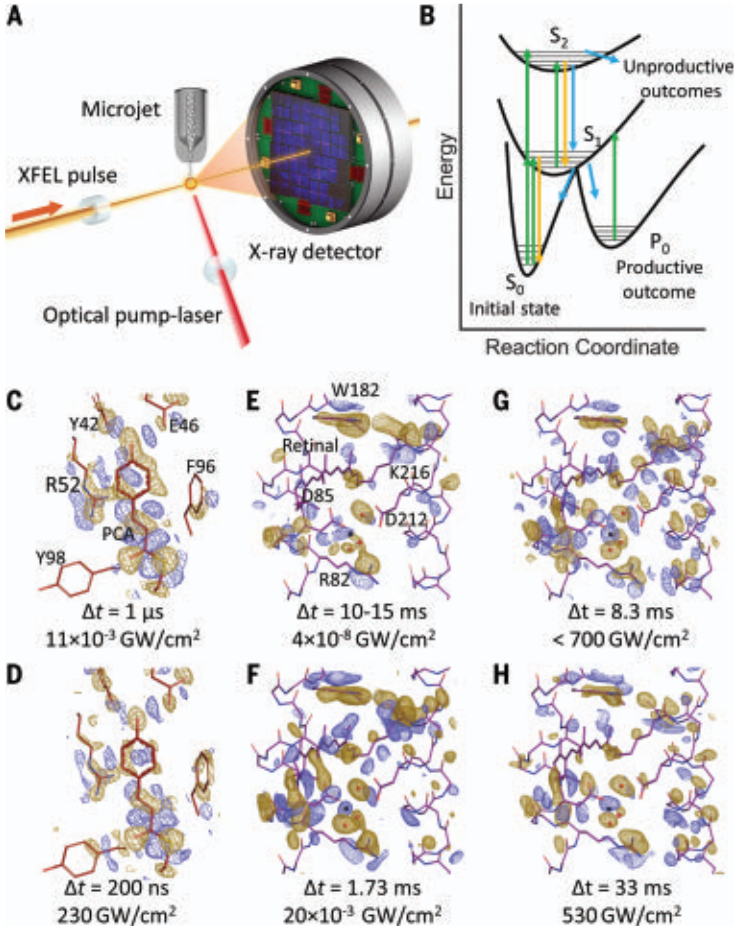


Fig. 2. Photoexcitation and multiphoton pathways. (**A**) Schematic illustrating a TR-SFX pump-probe study of a light-sensitive protein. Microcrystals are injected across a focused XFEL beam, and an optical pump laser photoexcites these microcrystals at a time delay Δt before the arrival of an XFEL pulse. X-ray diffraction data are recorded on an x-ray detector that is read out for each and every XFEL pulse. (**B**) Oversimplified scheme of photochemical pathways that may be populated at high-laser power densities. These include single-photon excitation of the ground state ($S_0 + h\nu \rightarrow S_1$, green arrow); coherent multiphoton activation of the ground state ($S_0 + 2h\nu \rightarrow S_2$, green arrows); excited state activation ($S_1 + h\nu \rightarrow S_2$, green arrow); photoreversal of the desired product ($P_0 + h\nu \rightarrow S_1$, green arrow); stimulated emission ($S_1 + h\nu \rightarrow S_0 + 2\,h\nu$ or $S_2 + h\nu \rightarrow S_1 + 2\,h\nu$, orange arrows); and various relaxation pathways ($S_1 \rightarrow P_0$, $S_1 \rightarrow S_0$, and $S_2 \rightarrow S_1$, blue arrows) including branching onto unproductive pathways (e.g., $S_2 \rightarrow$ U). Several other possible pathways are not illustrated (*49*). (**C** to **H**) Dependence of experimental isomorphous F_{obs}(activated) − F_{obs}(reference) difference Fourier electron density maps on the pump laser power density. 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. (C) Difference Fourier electron density map (positive electron density, blue; negative electron density, yellow; contoured at 3.0σ) calculated from PYP for the time point Δt = 1 μs when using a pump laser power density of 11×10^{-3} GW/cm^2 (Table 1) (*7*). PCA, *para*-coumaric acid. (D) Difference Fourier electron density map (contoured at 3.0σ) calculated from PYP for the time point Δt = 200 ns when using a pump laser power density of 230 GW/cm^2 (Table 1) (*61*). (E) Difference Fourier electron density map (contoured at 3.2σ) calculated from bR for the time point Δt = 10 to 15 ms when using a pump laser power density of 4×10^{-8} GW/cm^2 (Table 1) (*71*). (F) Difference Fourier electron density map (contoured at 3.7σ) calculated from bR for the time point Δt = 1.73 ms when using a pump laser power density of 20×10^{-3} GW/cm^2 (Table 1) (*70*). (G) Difference Fourier electron density map (contoured at 3.3σ) calculated from bR for the time point Δt = 8.3 ms when using a pump laser power density of <700 GW/cm^2 (Table 1) (*68*). For this experiment, the microcrystals were translated ~175 μm during the 8.33-ms time delay, and therefore the pump laser power density at the microcrystal position is considerably less than this nominal value but not known precisely. (H) Difference Fourier electron density map (contoured at 3.2σ) calculated from bR for the time point Δt = 33 ms when using a pump laser power density of 530 GW/cm^2 (Table 1) (*54*). The average power density within the full width at half maximum (FWHM) of the pump laser is quoted from Table 1.

Table 1. Experimental parameters for time-resolved x-ray diffraction studies of light-sensitive proteins. Mean fluence (F) and average power density through FWHM are quoted to approximately correct for x-ray and laser beam misalignments and spatial drift. $\sigma = \varepsilon \times 1000 \times \ln(10)/N_A$, where ε is the molar absorption coefficient and N_A is Avogadro's number. If $\sigma \cdot F/h\nu$ approaches or is larger than unity, the effective nonlinear cross sections of all accessible states must be used to estimate the number of absorbed photons, rather than using the resting-state cross section alone.

Publication	Pump laser energy (μJ)	Pump laser wavelength (nm)	Focal diameter FWHM* (μm)	Mean fluence through FWHM (mJ/cm^2)	Pump laser pulse duration (s)	Peak laser power (GW/cm^2)	Average power through FWHM (GW/cm^2)	Molar absorption coefficient† ($M^{-1}\cdot cm^{-1}$)	$\sigma \cdot F/h\nu$ on FWHM
Myoglobin									
Srajer *et al*. 1996 (*10*)	13,000	635	750	1470	7.5×10^{-9}	0.27	0.20	300	5.4
Schotte *et al*. 2003 (*11*)	23	570	200 × 120‡	61	1.0×10^{-12}	85	61	8,000	5.3
Barends *et al*. 2015 (*59*)	5	532	88	41	150×10^{-15}	377	272	12,000	5.0
Hemoglobin									
Knapp *et al*. 2006 (*16*)	1,100§	615	600	195	7×10^{-9}	0.039	0.028	600	1.4
Schotte *et al*. 2013 (*17*)	42	520	90 × 250‡	119	100×10^{-12}	1.6	1.2	31,000	37
PYP									
Genick *et al*. 1997 (*12*)	–	497	1000	2000	0.2	1.4×10^{-8}	1.0×10^{-8}	1,400	27
Perman *et al*. 1998 (*13*)	4,000	495	1500	113	7×10^{-9}	0.022	0.016	1,400	1.5
Schotte *et al*. 2012 (*14*)	–	390	120 × 600‡	350	120×10^{-12}	4.0	2.9	12,500	33
Jung *et al*. 2013 (*15*)	40	418	120 × 240‡	88	100×10^{-12}	1.2	0.9	30,000	21
Tenboer *et al*. 2014 (*7*)	15	485	150¶	76	7×10^{-9}	0.011	0.011	4,800	3.4
Pande *et al*. 2016 (*61*)	10	450	140	32	140×10^{-15}	322	232	42,000	12
Pandey *et al*. 2020 (*62*)	2.3	420	42	83	240×10^{-15}	479	345	27,500	20
FixL									
Key *et al*. 2007 (*18*)	1,000	615	1000	64	7.5×10^{-9}	0.012	0.009	1,500	1.1
Bacteriorhodopsin									
Nango *et al*. 2016 (*70*)	2 × 2§	532	40	80	4×10^{-9}	0.03	0.02	45,600	37
Nogly *et al*. 2018 (*68*)	17	529	56	70#	100×10^{-15}	4800	700#	45,600	32
Nass Kovacs *et al*. 2019 (*54*)	5.9	532	58	76**	145×10^{-15}	1060	530**	45,600	36
Weinert *et al*. 2019 (*71*)	14	520	102 × 61‡	143	5×10^{-3}	4×10^{-8}	4×10^{-8}	45,600	67
Other rhodopsins									
Skopintsev *et al*. 2020 (*72*)	3	575	47	86	150×10^{-15}	796	574	25,000	24
Yun *et al*. 2021 (*73*)	–	550	150	90	100×10^{-15}	1247	900	40,000	38
Oda *et al*. 2021 (*74*)	2 × 10§	470	40	386††	6×10^{-9}	0.09	0.06††	40,000	139
Switchable GFP									
Coquelle *et al*. 2018 (*63*)	1.5	400	40	94	230×10^{-15}	565	407	6,800	4.9
Woodhouse *et al*. 2020 (*65*)	46	400	180	90	100×10^{-12}	1.25	0.90	6,800	4.7
Phytochromes									
Claesson *et al*. 2020‡‡ (*64*)	40	640	100 × 80‡	318	70×10^{-15}	6300	4550	27,700	109
Carrillo *et al*. 2021‡‡ (*67*)	2 × 11§	640	52	260	5×10^{-9}	0.07	0.05	27,700	88
Photosystem II									
Suga *et al*. 2017 (*79*)	2 × 19§	532	240¶	42	7×10^{-9}	0.006	0.006	5,000	2.1
Kern *et al*. 2018 (*80*)	–	527	–	120	120×10^{-9}	0.0014	0.0010	5,000	6.0
Suga *et al*. 2019 (*81*)	19,500	532	5500¶	82	7×10^{-9}	0.012	0.012	5,000	4.2
Ibrahim *et al*. 2020 (*82*)	–	532; 527	–	120	8; 120×10^{-9}	0.021	0.015	5,000	6.0
Reaction center									
Baxter *et al*. 2004 (*20*)	3,000	630	1000	190	7×10^{-9}	0.038	0.027	25,000	58
Wöhri *et al*. 2010 (*19*)	200	527	200 × 600‡	106	100×10^{-9}	0.0015	0.0011	100,000	107
Dods *et al*. 2021 (*83*)	11.8	960	190	21	150×10^{-15}	192	139	100,000	38
Cytochrome c oxidase									
Shimada *et al*. 2017 (*96*)	100§	532	200	159	4×10^{-9}	0.055	0.040	16,500	27

*When it was not specified whether the focal spot diameter corresponded to $1/e^2$ or to FWHM, we assumed FWHM. †Estimated from literature sources and may deviate from values in the crystalline form. ‡Ellipsoidal spot profile. §Incoming laser split into two beams from opposite sides. Values correspond to the power density and fluence for only one of these beams. ¶Top-hat profile for the laser with fluence within the FWHM given in the publication. All other profiles are treated as Gaussian. #Focal spot offset 50 μm from the laser peak maximum so that the fluence at the crystal position does not correspond to maximum fluence. **Focal spot offset 25 μm from the laser peak maximum so that the fluence at the crystal position does not correspond to maximum fluence. ††Focal spot offset 6 μm from the laser peak maximum so that the fluence at the crystal position does not correspond to maximum fluence. ‡‡When the grease carrier medium was mixed with crystals and buffer, this attenuated the light intensity by up to two orders of magnitude.

benchmark light-induced electron density changes measured using XFEL radiation (*7*) against earlier time-resolved Laue diffraction studies. PYP is a blue-light receptor that contains a buried *para*-coumaric acid chromophore that is photoisomerized to initiate a structural transition to the protein's signaling state. Convincing difference Fourier electron density maps (*50*) (Box 3) presented the experimental differences between data collected with and without photoactivation as changes in electron density for the time delays Δt = 10 ns and 1 μs (*7*) (Fig. 2C). The lower optical density of microcrystals allowed a higher fraction of photoexcited PYP to be achieved when compared with Laue diffraction studies using larger crystals (*12*–*15*), and this approach improved the quality of the observed electron density changes (*7*). Whereas experimental fluctuations between individual diffraction intensity measurements were large, such noise could be averaged out by merging diffraction data from a very large number of microcrystals. In practice, 10,000 to 50,000 processed diffraction images are usually required to recover high-quality difference Fourier electron density maps.

Ultrafast motions in biological structures

Light-sensitive proteins contain various chromophores with conjugated double and/or aromatic bonds. After a photon is absorbed, the committed step—such as the isomerization of a chromophore, the movement of an electron, or the dissociation of a ligand—is typically complete within hundreds of femtoseconds to a few picoseconds. Protein conformational changes are not hindered by thermal energy barriers on this time scale, and therefore ultrafast trajectories cannot be visualized using low-temperature trapping methods (*51*). XFEL radiation thus creates an experimental opportunity to elucidate the influence of structural changes on ultrafast biological reactions and draw comparison with theoretical predictions (*52*). Nonetheless, the requirement for high crystallographic occupancy in time-resolved diffraction studies (Box 3) led to the first ultrafast TR-SFX experiments being performed using pump laser power densities (Table 1) that would be regarded as excessive in time-resolved spectroscopy applications. Consequently, the biological relevance of structural observations recovered when working with power densities a few tens or hundreds of gigawatts per square centimeter has been debated (*53*–*55*) (Box 4).

Ultrafast protein breathing motions were observed in XFEL-based TR-XSS studies of a photosynthetic reaction center under artificial conditions of extreme multiphoton excitation (*56*) (Table 2 and Movie 1). TR-XSS experiments also revealed underdamped oscillatory motions with a period of 3.6 ps after CO was photodissociated from the heme cofactor of myoglobin (*57*), although this signal may have arisen from density modulations within the solvent surrounding the protein rather than from oscillations within the protein (*58*). Considerably more structural details were provided by TR-SFX studies of Mb:CO (*59*), in which correlated movements of α helices A and E toward the heme and α helices C, F, and H away from the heme were observed. Although the biological function of myoglobin is to assist the diffusion of oxygen within muscles, the photodissociation of CO from myoglobin is a heavily studied reaction that has provided structural insight into ultrafast protein motions associated with the breaking of a metal-ligand bond. Heme doming, in which the iron moves out of the plane of the protoporphyrin IX ring as the ring periphery moves in the opposite direction (Fig. 3A), arose faster than the time resolution of this study (250 fs) (*59*). This finding is consistent with XFEL-based time-resolved x-ray absorption spectroscopy studies of Mb:CO for which a rapid signal (~70 fs) was assigned to doming of the heme iron and a slower component (~400 fs) was attributed to a displacement of helix F (*60*).

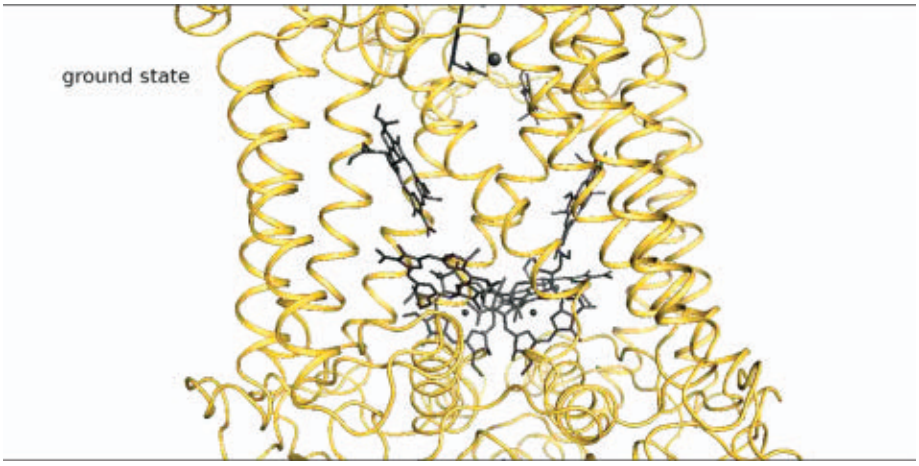
TR-SFX has also enabled ultrafast isomerization trajectories within naturally light-sensitive proteins to be visualized, including PYP (*61*, *62*), a photoswitchable green fluorescent protein (rsEGFP2) (*63*), and a bacterial phytochrome (*64*). TR-SFX studies of PYP revealed a structural transition at $\Delta t \approx 550$ fs associated with the trans-to-cis transition of its *para*-coumaric acid chromophore. This reaction was modeled as the isomerizable double bond first adopting a highly twisted trans configuration before rotating a further 90° to reach

Box 4. Photoexcitation and multiphoton excitation.

Time-resolved diffraction studies of light-sensitive proteins aim to visualize how the absorption of a single photon induces biologically important conformational changes. In an idealized scenario, a pump laser pulse will photoactivate a high fraction of the protein molecules within a crystal (Fig. 2A) from their resting state to the first excited state (Fig. 2B), which evolves along the reaction coordinate before decaying into the ground state of the biologically relevant trajectory. The product of the protein's resting-state absorption cross section (σ) with the pump laser fluence (F) divided by the photon's energy ($h\nu$) provides an estimate of the number of photons absorbed per protein molecule when $\sigma \cdot F/h\nu \ll 1$. When $\sigma \cdot F/h\nu$ is the order of unity or greater (Tables 1 and 2), and especially when using a pump laser pulse with a duration of $\leq 10^{-13}$ s, coherent multiphoton activation, stimulated emission, stimulated Raman scattering, branching onto unproductive pathways, photoreversal of the desired product (Fig. 2B), and numerous other reactions (*53*–*55*, *127*, *128*) may all be induced to a considerable extent. Under these conditions, the photoexcitation pathways are determined by the effective nonlinear cross sections of all accessible states (*55*, *127*) rather than the ground-state cross section alone. A microjet may also focus the pump laser beam (*129*), and estimates for losses due to light scattering from and within a microjet diverge markedly (*54*, *56*, *64*, *67*, *68*, *129*). Consequently, the product $\sigma \cdot F/h\nu$ systematically overestimates the number of photons absorbed per protein molecule when compared with values extracted from TR-XSS data (Table 2).

Photoexcitation of competing reaction pathways risks obscuring the biological interpretation of any measured structural change (*53*). Difference electron density arising from valence electron changes is not expected to be visible below 0.8-$Å^{-1}$ resolution (*49*), and therefore the photoexcitation of aromatic residues and multiphoton ionization processes (*53*, *54*) have not been visualized directly in any TR-SFX study to date. Moreover, because x-ray diffraction averages scattering from billions of molecules within each crystal, time-resolved diffraction cannot usually resolve electron density changes arising from conformations with much lower occupancy than the prevailing structural trajectory. Multiphoton excitation, however, may perturb reaction pathways by altering the distribution of charges or increasing the kinetic energy of light-absorbing cofactors. In one extreme case, the absorption of ~800 photons per molecule (Table 2) induced a quasi-isotropic expansion of a bacterial photosynthetic reaction center (*56*), with Cα atoms displaced by up to 0.5 Å within 7 ps (Movie 1). Despite these concerns, TR-SFX experiments spanning 10 orders of magnitude in pump laser power density (Table 1) show electron density changes that correlate with functional pathways and concur with biologically important structural changes on microsecond (Fig. 2, C and D) (*7*, *62*) and millisecond (Fig. 2E-H) (*54*, *68*, *70*, *71*) time scales. These observations suggest that proteins may have evolved to direct all deposited energy toward functional outcomes (*53*). Although these observations do not prove that multiphoton and single-photon pathways are identical, they highlight the value of working above single-photon excitation thresholds (Table 1) to maximize crystallographic occupancy (Box 3) while remaining within the regime for which the evolution of spectral changes is consistent with single-photon excitation. It is, however, important to repeat ultrafast TR-SFX studies at lower power densities once photoexcitation conditions have been optimized using time-resolved spectroscopy (*55*, *127*).

a twisted cis conformation by $\Delta t \approx 700$ fs (*61*) (Fig. 3B and Movie 2). Similarly, TR-SFX studies of the light-triggered switch from a nonfluorescent to a fluorescent conformation of the hydroxybenzylidene imidazolinone chromophore of rsEGFP2 showed that by $\Delta t = 1$ ps, the photoisomerization reaction had proceeded to a twisted configuration about halfway between the cis (nonfluorescent) and trans (fluorescent) isomers (*63*) (Fig. 3C) and was fully isomerized by $\Delta t = 10$ ns (*65*). The protein binding pocket's ability to accommodate the twisted conformation was argued to be essential for photoswitching. A twisted conformation also characterized the photoisomerization trajectory of the biliverdin chromophore of bacterial phytochromes. By $\Delta t = 1$ ps, the C15=C16 double bond linking the C and D rings of biliverdin had rotated ~50° (Fig. 3D) (*64*), which was both considerably faster and in the opposite direction to what had been anticipated (*66*). Prominent difference electron density features were observed to have spread entirely throughout a related phytochrome by $\Delta t = 5$ ns (*67*), as the structural signal from biliverdin isomerization is rapidly propagated.



Movie 1. Ultrafast structural response of a bacterial photosynthetic reaction center to extreme multiphoton excitation as measured using TR-XSS. These motions reached a maximum amplitude at $\Delta t = 7$ ps and decayed with a half-life of 44 ps (*56*).

Structural trajectories of microbial rhodopsins

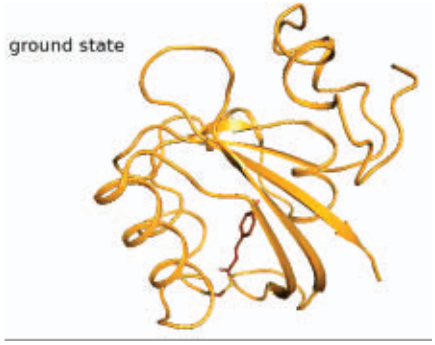
Microbial rhodopsins are a family of seven-transmembrane, retinal-binding proteins including light-driven ion pumps, light-gated ion channels, and light sensors. Bacteriorhodopsin (bR) is an archetypal proton pump and, in its resting conformation, it contains an all-trans 15-anti retinal covalently bound to Lys^{216} through a protonated Schiff base. Retinal was observed to transition through a twisted geometry at

Table 2. Experimental parameters for time-resolved x-ray solution scattering studies of light-sensitive proteins. Photon absorption (rightmost column) calculated from the measured temperature change multiplied by the heat capacity divided by the energy per photon divided by the protein concentration. Heat capacities were approximated as 4.1 J/cm^3 for soluble protein solutions and 3.8 J/cm^3 for membrane protein solutions containing detergent. TR-XSS measurements are not sensitive to multiphoton pathways for which a large fraction of the photon's energy does not ultimately emerge as heat, such as stimulated emission or stimulated Raman scattering. Energy stored within the protein will also not be visible as heat.

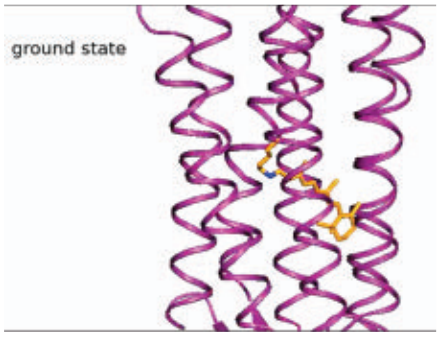
Publication	Pump energy (μJ)	Pump wavelength (nm)	Focal diameter FWHM (μm)	Mean fluence through FWHM (mJ/cm^2)	Pulse duration (s)	Peak laser power (GW/cm^2)	Average power through FWHM (GW/cm^2)	Molar absorption coefficient* (M^{-1}·cm^{-1})	[Protein] (mM)	ΔT (°C) measured	$\sigma \cdot F/h\nu$ on FWHM	Photons absorbed per protein molecule
Hemoglobin												
Cammarata *et al.* 2008 (*22*)	500	527	1000 × 220†	145	150×10^{-9}	0.0013	0.0010	40,000	1.2	1.8	59	27
Myoglobin												
Cho *et al.* 2010 (*24*)	115	480	120 × 600†	102	35×10^{-12}	4.0	2.9	7,000	3.0	0.35	6.6	1.9
Levantino *et al.* 2015 (*57*)	25	538	300	18	250×10^{-15}	98	71	14,500	2.35	0.10‡	2.7	0.8
PYP												
Kim *et al.* 2016 (*25*)	–	460	–	100	7×10^{-9}	0.020	0.014	26,000	4.4	0.40	23	1.4
Bacteriorhodopsin												
Andersson *et al.* 2009 (*23*)	250	527	265	227	4×10^{-9}	0.079	0.057	45,600	0.57	0.50	105	15
DrBphP												
Björling *et al.* 2016 (*26*)	–	660	–	200	5×10^{-9}	0.056	0.040	38,000	0.3	0.50	96	38
Björling *et al.* 2016 (*26*)	–	680	–	200	2×10^{-12}	139	100	54,000	0.3	0.14	141	11
RCvir												
Arnlund *et al.* 2014 (*56*)	260	800	147	764	500×10^{-15}	2120	1530	180,000	0.22	7.0	1180	808§

*Estimated from literature sources. †Ellipsoidal spot profile. ‡An upper estimate for the induced heating, given that the agreement with heating calibration data recorded at a synchrotron was only approximate. §Extreme case of coherent multiphoton excitation for which the appropriate cross section has not yet been reported.

Movie 2. Ultrafast structural response of PYP to the photoisomerization of its *para*-coumaric acid chromophore measured using TR-SFX. This movie is constructed from the sequences of PDB IDs 5HD3 (resting), 5HDC (Δt = 100 to 400 fs), 5HDD (Δt = 800 to 1200 fs), and 5HDS (Δt = 3 ps).



Movie 3. Ultrafast structural response of bR to the photoisomerization of its retinal chromophore measured using TR-SFX. This movie is constructed from the sequences of PDB IDs 6G7H (resting), 6G7I (Δt = 49 to 406 fs), 6G7J (Δt = 457 to 646 fs), and 6G7K (Δt = 10 ps).



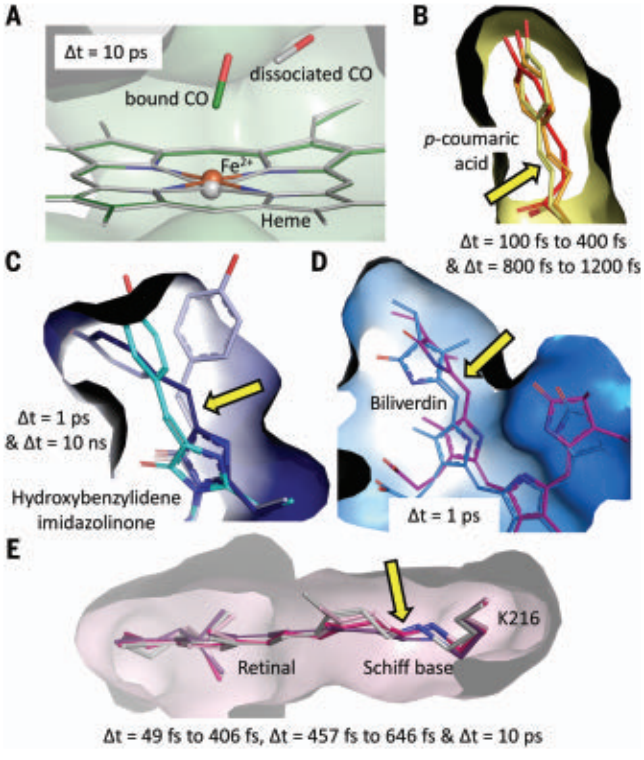


Fig. 3. Ultrafast distortions of protein chromophores. (**A**) Heme doming associated with the photodissociation of CO from the active site of myoglobin (*59*). Superimposed on the resting conformation [green, Protein Data Bank (PDB) ID 5CMV] is the refined structure for Δt = 10 ps (light gray, PDB ID 5CNE). (**B**) Twisting of the *para*-coumaric acid chromophore of PYP during photoisomerization about the C2=C3 bond (*61*). Superimposed on the resting conformation (yellow, PDB ID 5HD3) are refined structures for Δt = 100 to 400 fs (orange, PDB ID 5HDC) and Δt = 800 to 1200 fs (red, PDB ID 5HDD). (**C**) Twisting of the hydroxybenzylidene imidazolinone chromophore of rsEGFP2 during photoisomerization about the CA2=CB2 bond (*63*). Superimposed on the resting conformation (dark blue, PDB ID 5O8A) are refined structures for Δt = 1 ps (cyan, PDB ID 5O8B) and Δt = 10 ns (light blue, PDB ID 6T3A). (**D**) Twisting of the biliverdin chromophore of a bacterial phytochrome during photoisomerization about the C15=C16 bond (*64*). Superimposed on the resting conformation (blue, PDB ID 6T3L) is the refined structure for Δt = 1 ps (purple, PDB ID 6T3U). (**E**) Twisting of the retinal chromophore of bR during photoisomerization about the 13=14 bond (*68*). Superimposed on the resting conformation (dark purple, PDBID 6G7H) are refined structures for Δt = 49 to 406 fs (violet, PDB ID 6G7I) and Δt = 457 to 646 fs (light purple, PDBID 6G7J) and Δt = 10 ps (gray, PDB ID 6G7K). Yellow arrows in (B) to (E) indicate the photoisomerized bond. Surfaces represent the protein's pocket surrounding the chromophore.

$\Delta t \approx 450$ fs about halfway between the trans and cis configurations, before relaxing into a twisted 13-cis geometry on the picosecond time scale (*54*, *68*) (Fig. 3E). Low-amplitude α helix motions arose and decayed as retinal isomerized (*69*) (Movie 3), illustrating the subtle, yet surprisingly complex, nature of protein dynamics on this time scale. These motions enhanced the isomerization yield by steering the position of a water molecule (w_{402}, Fig. 4, Δt = 49 to 406 fs), which forms an H-bond with the Schiff base nitrogen in the bR resting state, thereby guiding the retinal through its isomerization trajectory (*68*) and into the K state (Fig. 4). Vibrational motions with a period of ~400 fs were also observed for residues and water molecules in the immediate surroundings of the retinal chromophore (*54*), illustrating how vibronic coupling may also enhance the isomerization yield.

Once retinal has isomerized, the next key event is the L-to-M transition (Fig. 4) resulting from a proton being transferred from the protonated Schiff base to Asp^{85}, a reaction that occurs on a time scale of microseconds. One long-standing puzzle is why retinal photoisomerization rotates the Schiff base proton toward the cytoplasm when protons are released into the extracellular medium. TR-SFX studies of bR (*70*) revealed that w_{402} disorders within a few nanoseconds after retinal isomerization (Fig. 4, Δt = 760 ns), which disrupts the extended extracellular H-bond network of charged residues and water molecules that separate the proton affinities of the donor (the Schiff base) and acceptor (Asp^{85}) by 11 orders of magnitude. A rearranged network of H-bond interactions then evolves within microseconds, which raises the pK_a (where K_a is the acid dissociation constant) of Asp^{85}, lowers the pK_a of the Schiff base, and establishes a pathway for proton exchange between these two groups via a transiently ordering water molecule (w_{452}) and Thr^{89} (Fig. 4, Δt = 760 ns). The futile reverse proton transfer from Asp^{85} to the Schiff base is prevented by the H-bond connection between Asp^{85} and Thr^{89} being broken once Asp^{85} is protonated (*70*) and other structural changes ensure that a proton is released to the extracellular medium. Time-resolved serial synchrotron crystallography (TR-SSX) studies of bR revealed the transient ordering of three more water molecules (Fig. 4, Δt = 10 to 15 ms) and the opening of the protein toward the cytoplasm (*71*). These structural changes create a pathway for Schiff base reprotonation from the cytoplasm and complete the proton pumping cycle.

Light-induced electron density changes have recently been reported for three other microbial rhodopsins: sodium (KR2) (*72*) and chloride (CIR) (*73*) pumps originating from marine flavobacteria and a chimeric construct of the

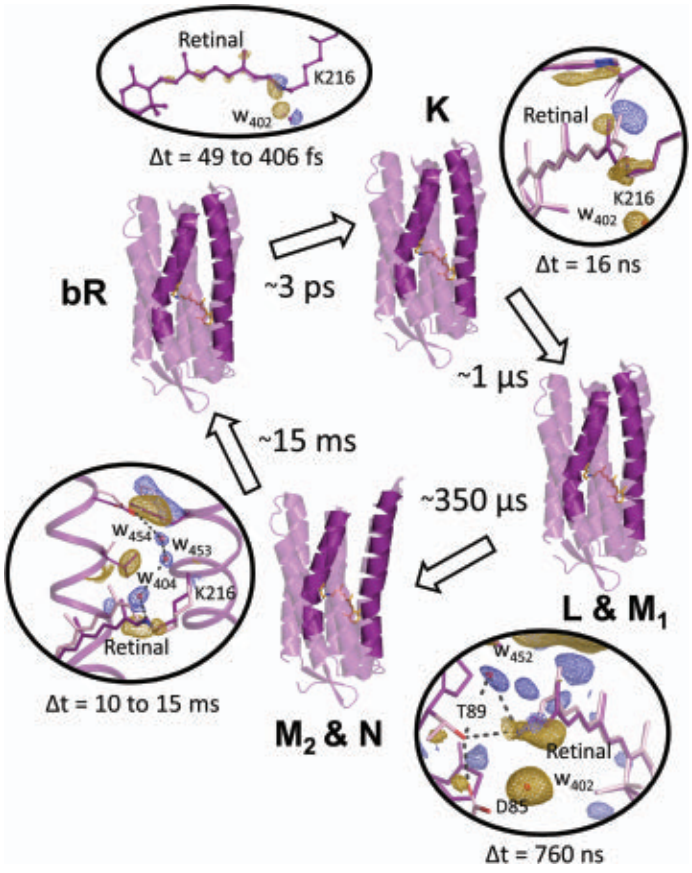


Fig. 4. Light-initiated structural changes driving proton pumping in bR. Protein conformations associated with selected spectral states (bR resting state, K state, L and M_1 states, and M_2 and N states) are arranged as a photocycle with an approximate time scale for each structural transition indicated. Refined light-activated structures (light purple), the resting bR structure (dark purple), and the corresponding isomorphous F_{obs}(activated) – F_{obs}(reference) difference Fourier electron density map (positive electron density, blue; negative electron density, yellow) are shown as insets. For Δt = 49 to 406 fs, the map is calculated from data associated with PDB ID 6G7I minus data associated with PDB ID 6G7H and is contoured at ±4.6σ (where σ is the root mean square electron density of the unit cell). For Δt = 16 ns, the map is calculated from PDB IDs 5B6W minus 5B6V and is contoured at ±3.7σ. For Δt = 760 ns, the map is calculated from PDB IDs 5B6X minus 5B6V and is contoured at ±3.1σ. For Δt = 10 to 15 ms, the map is calculated from PDB IDs 6RPH minus 6RQP and is contoured at ± 3.0σ.

light-gated channelrhodopsins 1 and 2 (C1C2) deriving from green alga (*74*). Picosecond TR-SFX data showed that retinal had isomerized to a 13-cis conformation within 2 ps of photoactivation in both CIR (*74*) and KR2 (*72*), with the retinal's C20 methyl group displaced toward helix C in both cases. This contrasts with bR, for which the C20 methyl initially tilts toward helix G (*68*, *70*) and illustrates how strain energy stored in a nonplanar retinal is adapted to protein functionality. A chloride ion bound in the local vicinity of the retinal in CIR was also displaced within a few picoseconds (*73*), analogous to the rapid displacement of w_{402} observed in bR (*54*, *70*). TR-SFX data recorded for KR2 (*72*) and C1C2 (*74*) on a microsecond to millisecond time scale showed retinal isomerization–induced conformational changes associated with helix C, but larger amplitude motions of the cytoplasmic portions of helices E and F (*71*) did not arise. Side-chain rearrangements in KR2 created two putative sodium-binding sites on the extracellular side of the retinal which, in combination with Schiff base deprotonation neutralizing a positive electrostatic barrier, were suggested to facilitate the active transport of sodium out of the cell (*72*). By contrast, in C1C2, a prominent outward displacement of helix C near the retinal binding pocket (*74*) was hypothesized to open a constriction site and thereby facilitate the passive permeation of ions across a biological membrane. Altogether, these structural findings illustrate how a shared transmembrane heptahelical scaffold uses diverse reaction pathways to achieve distinct biological tasks.

Structural changes during the water-splitting reaction of photosynthesis

Plants, cyanobacteria, and photosynthetic bacteria use integral membrane protein complexes to convert the energy of a captured photon into the movement of an electron. A conserved structural core at the heart of these complexes includes 10 transmembrane helices that support several photosynthetic pigments arranged around an axis of pseudo-twofold symmetry. Dioxygen (O_2) production at the oxygen-evolving complex (OEC) of photosystem II (PSII) (Fig. 5A) of plants and cyanobacteria is one of the most intensely studied biochemical reactions. TR-SFX has recently allowed structural changes in cyanobacterial PSII to be visualized (*75*–*82*). In these experiments, a laser flash photooxidizes a special pair of closely spaced chlorophyll molecules (P_{680}) near the OEC, and the liberated electron reduces a mobile plastoquinone molecule located ~3 nm away and on the opposite side of the membrane. TR-SFX studies of a purple bacteria photosynthetic reaction center (which lacks the OEC) reveal that the equivalent light-driven charge separation reaction induces a subtle adjustment of the protein's structure in response to the movement of an electron and that these rearrangements, occurring on a picosecond time scale, help to stabilize the separated charges (*83*).

In PSII, the OEC supplies the electrons needed to reduce photooxidized P_{680}^+ and comprises four manganese atoms, one calcium atom, and five structural water molecules (Mn_4CaO_5) (*84*). According to Kok's S-state clock (*85*), four photoactivation events advance the OEC through a sequence of metastable intermediates called S_1, S_2, and S_3 and the transient S_4 state, which spontaneously decays to S_0 (*85*) as molecular oxygen generated by the water-splitting reaction $2H_2O \rightarrow O_2 + 4e^- + 4H^+$ is released from the OEC. PSII microcrystals are prepared in the S_1 state (*75*–*81*), and the S state probed is determined from the number of laser flashes each microcrystal experiences, with a recommended delay of 100 ms or more between each laser flash to accommodate the reaction kinetics of the plastoquinone electron acceptor (*86*). Proton release and uptake events on opposite sides of an energy-transducing biological membrane result in four protons being transferred from the stroma to the lumen for every O_2 molecule produced. Structural studies of PSII have also driven innovation in sample-delivery technologies, including simultaneous x-ray emission spectroscopy and x-ray diffraction measurements from microcrystals (*75*), the development of drop-on-tape sample-delivery systems (*80*, *87*), and flash-freeze approaches that use microcrystals mounted on a solid mesh (*81*).

One key feature of femtosecond XFEL pulses when studying PSII is that they do not allow time for x-ray–induced radiation damage (*88*) to cause the photoreduction of metal centers (*89*). X-ray emission spectroscopy data collected from PSII microcrystals confirmed that the Mn_4CaO_5 complex is not photoreduced by the XFEL beam (*75*). A high-resolution structure of PSII recovered by merging XFEL diffraction data collected from large crystals cooled to 100 K (*90*) showed subtle differences in the interatomic distances (~0.1 to 0.2 Å) of the OEC relative to a similar resolution structure solved using synchrotron radiation (*84*). Although these differences may arise from the OEC being photoreduced when using synchrotron radiation, coordinate errors also play

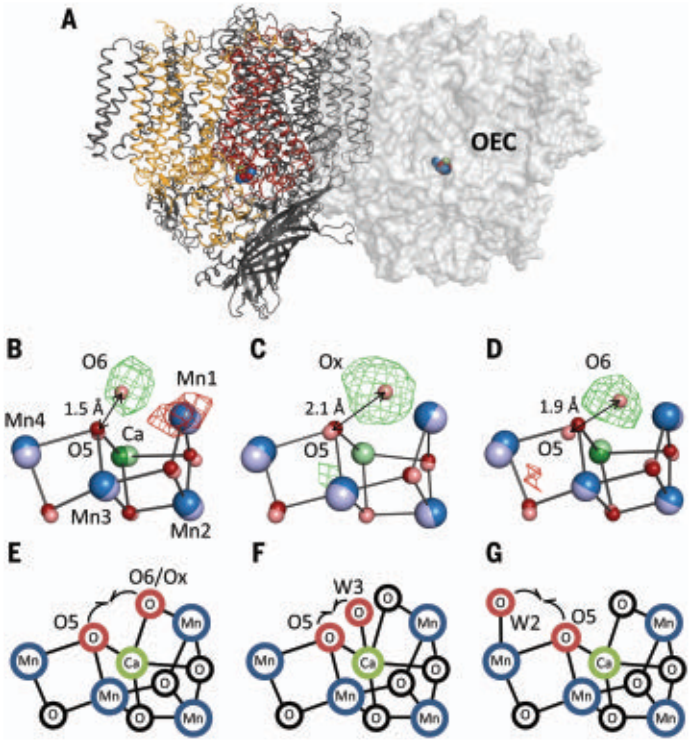


Fig. 5. Electron density changes during the water splitting–catalyzed reactions of PSII. (**A**) Overview showing the structure of PSII and location of the OEC. (**B**) $F_{obs} - F_{calc}$ omit electron density map (PDB ID 5WS6, calculated with O6 removed from the model) associated with the S_3 state of PSII. Data were recorded 10 ms after two laser flashes separated by 10 ms. (**C**) $F_{obs} - F_{calc}$ omit map (PDB ID 6DHO, calculated with Ox removed from the model) associated with the S_3 state of PSII. Data were recorded 200 ms after two laser flashes separated by 200 ms. (**D**) $F_{obs} - F_{calc}$ omit map (PDB ID 6JLL calculated with O6 removed from the model) associated with the S_3 state of PSII. Data were recorded from microcrystal immediately flash-frozen after two laser flashes separated by 1 s. Green (red) mesh shows positive (negative) $F_{obs} - F_{calc}$ omit electron density contoured at ±3.5σ. (**E**) Schematic indicating a putative mechanism whereby O5 and O6 (Ox) combine to form O_2 (*79*–*81*). (**F** and **G**) Schematic indicating a putative mechanism whereby O5 and a water molecule, either W3 (F) which coordinates Ca^{2+} of the OEC or W2 (G) which coordinates Mn4 of the OEC, combine to form O_2 (*80*). Substrate oxygen atoms are indicated in red, Mn atoms in blue, Ca^{2+} in green, and other oxygen atoms in black. For simplicity, only the atoms of the OEC or ligands participating in O_2 formation are shown. This scheme is adapted from that given in reference (*80*).

a role, as all high-resolution structures of PSII solved using XFEL radiation (*78*–*82*) show similar variations in their internal OEC distances.

Initial TR-SFX studies of PSII were at low (~4.5-Å) resolution (*75*–*77*). Two of these studies did not report any discernible conformational changes associated with S_2 (*75*) or S_3 (*76*) relative to S_1. A third study argued that the Mn_4CaO_5 complex is elongated in S_3 and a loop containing OEC ligands is displaced (*77*), but these conclusions were contested (*86*). A breakthrough came when changes in the growth, handling, and the use of larger PSII microcrystals allowed TR-SFX diffraction data to be recorded at 2.35-Å resolution or better, at which point light-induced changes in electron density in the OEC could be resolved (*78*–*82*) and movements resulting from the reduction of plastoquinone observed (*79*, *80*, *82*).

In the resting S_1 state, the OEC has Mn1, Mn2, Mn3, Ca, and four oxygen atoms (O1, O2, O3, and O5) arranged as eight corners of a distorted cube, whereas Mn4 and O4 occupy positions defining the backrest of a distorted-chair configuration (*84*). X-ray structures show that Mn1 and O5 are separated by 2.7 Å, which is too long for a ligand bond, and the complex is said to adopt an open cubane geometry (*80*, *84*). Progression from S_1 to S_2 involves only subtle changes, and the coordination states of all OEC metals are unchanged (*80*–*82*). The S_2-to-S_3 transition, however, is characterized by the transient ordering of a new oxygen atom, named O6 (*79*, *81*) or Ox (*80*, *82*), which occupies the sixth coordination site of Mn1, provides a second bridge from Mn1 to Ca, and is in close proximity to O5. The distance separating this new oxygen atom and O5 was first modeled as 1.5 Å (*79*), which corresponds to a peroxide ($O\text{-}O^{2-}$) with O5 and O6 covalently bound (Fig. 5B). In a subsequent study, this distance was modeled as 2.1 Å, and the new oxygen atom was assigned as either a hydroxide (OH^-) or an oxo (O^{2-}) species (*80*) (Fig. 5C). A third study modeled this separation as 1.9 Å, suggesting an oxyl ($\cdot O^-$) oxo (O^{2-}) pair for O5 and O6, with the specific assignment dependent on the redox state of the Mn atoms to which they ligate (*81*) (Fig. 5D). Finally, the most recent study models this separation as 1.9 ± 0.14 Å 150 μs after a second laser flash, maturing to 2.2 ± 0.25 Å in the S_3 state. Moreover, the insertion of Ox was shown to correlate with the oxidation state of Mn1 (*82*). The transient S_4 state and its decay into the metastable S_0 conformation (*80*) were also visualized using TR-SFX, revealing that a new water molecule replaces the oxygen atom lost at the O5 position as O_2 is released. Movements of several residues, including D1-Tyr^{161} (or Tyr_Z) (*82*), and neighboring water molecules were also visible in these very rich TR-SFX data (*78*–*82*), suggesting pathways for the uptake of water and the release of protons from the OEC.

TR-SFX studies have thus converged on a model in which O5 is one substrate oxygen (*79*–*82*) that either combines with O6 (Ox) to form dioxygen (*79*–*81*) (Fig. 5E) or, alternatively, another water molecule attacks O5 (Fig. 5, F and G) and O6 (Ox) moves to occupy the vacated O5 position as dioxygen is released (*80*). A unique assignment of all oxygen species in the S_3 state is required to distinguish between these competing models (*82*). A consensus picture must also be consistent with the oxidation and spin states of each Mn atom of the OEC (*91*) and be supported by quantum chemistry calculations (*92*). Thus, the accuracy of the O-O and O-Mn distances is critical and highlights a weakness of time-resolved crystallography, given that modeling these distances with partial occupancy to an accuracy better than 0.1 Å is testing the limits of the method and judgment is needed when applying nonstandard bond-length constraints.

Dynamical studies of enzymes that are not light sensitive

Light-sensitive proteins are ideal targets for time-resolved diffraction owing to the ease with which the reaction of interest can be initiated, but they encompass only a small fraction of biologically important enzymes. Other approaches to reaction triggering are therefore essential for the field to grow to its full potential. Convincing time-dependent electron density changes have been reported for a smaller number of enzymes that are not naturally light sensitive when mixing crystals with a substrate to initiate a reaction or when using photocaged compounds or other strategies that allow light to be used as a trigger.

One such example is cytochrome c oxidase (CcO), which couples the four-electron reduction of dioxygen to water to proton pumping and thus catalyzes the opposite chemical reaction to PSII. As a metal-containing redox enzyme, CcO is sensitive to x-rays, and CcO crystals can be photoreduced when diffraction data are collected at a synchrotron radiation source. A damage-free structure of bovine

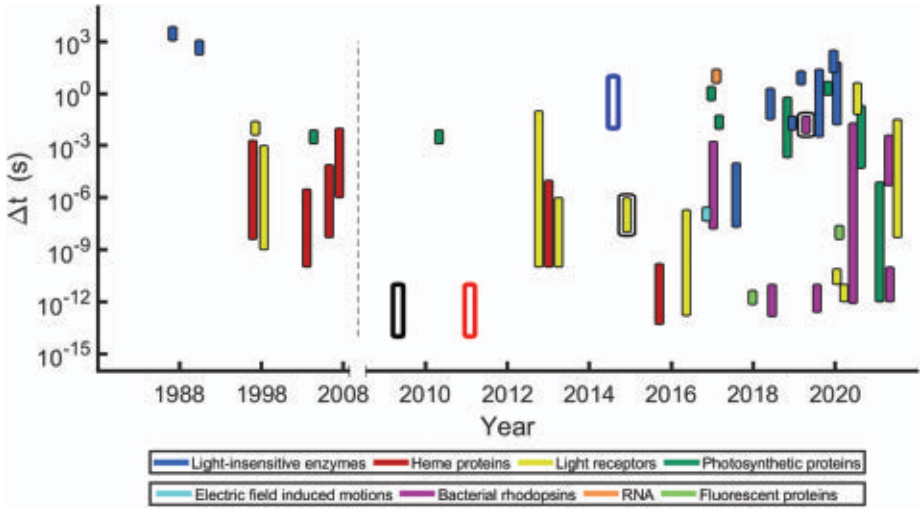


Fig. 6. Summary of time-resolved crystallography studies using synchrotron and XFEL radiation. The date of publication (*x* axis) and the time-domain of each study (*y* axis) are indicated. Only time-resolved diffraction studies reporting diffraction data beyond 4-Å resolution are represented. Low-temperature kinetic crystallography or intermediate trapping studies are not shown, bar one exception (*81*). Structural results are grouped as follows: studies of enzymes that are not naturally light sensitive [light-insensitive enzymes: glycogen phosphorylase B (*2*), Ha-Ras p21 (*3*), cytochrome c oxidase (*96*, *98*), fluoroacetate dehalogenase (*107*), xylose isomerase (*100*), β-lactamase (*101*), P450 nitric oxide reductase (*106*), and isocyanide hydratase (ICH) (*102*)]; studies of heme-containing proteins from which a CO ligand was photodissociated [heme proteins: myoglobin (*10*, *11*, *59*), dimeric hemoglobin (*16*), heme domain of FixL (*18*), and tetrameric hemoglobin (*17*)]; studies of light receptors [photoactive yellow protein (*7*, *12*–*15*, *61*, *62*), bacterial phytochromes (*64*, *67*), and LOV2 domain of phototropin-2 (*113*)]; studies of photosynthetic proteins [photosynthetic reaction center (*19*, *20*, *83*) and photosystem II (*78*–*82*)]; studies of electric field–induced motions [PDZ domain of LNX2 (*21*)]; studies of bacterial rhodopsins [bacteriorhodopsin (*54*, *68*, *70*, *71*), sodium pumping KR2 (*72*), chloride pumping CIR (*73*), and chimeric channelrhodopsin (*74*)]; studies of RNA [riboswitch RNA (*103*)]; and studies of fluorescent proteins [rsEGFP2 (*63*, *65*)]. Time-resolved diffraction studies that sampled a time domain overlapping with earlier work on the same protein are not shown unless they used a different approach, emphasized a different structural conclusion, or were associated with a technical milestone. Empty rounded rectangles represent the first XFEL operating at 1.5 Å (black), the first demonstration of SFX (*5*) (red), and the first demonstration of synchrotron serial crystallography (*130*) (blue). First demonstrations of TR-SFX (*7*) and TR-SSX (*71*) are indicated with black rounded rectangles around those experiments. Some adjustments to publication dates were made to minimize the overlap of bars.

CcO was first recovered by collecting data from large crystals cooled to cryogenic temperature using XFEL radiation (*93*). Room-temperature SFX structures of bovine CcO (*94*) and CcO from a thermophilic bacteria (*95*) followed shortly after, with electron density at the heme-copper active site being interpreted as either a hydroxide (or water) (*95*) or peroxide (*93*) species. As with other heme-containing enzymes (*10*, *11*, *16*–*18*), the first time-resolved diffraction study used large crystals of reduced CcO with CO bound to the active site heme, and this ligand was photodissociated using a laser flash (*96*). An innovative study—which aimed to observe structural intermediates of the dioxygen reduction reaction—used a mixing jet (*97*) to combine microcrystals of reduced CcO with a saturated oxygen solution (*98*). Spectroscopic measurements showed that the enzyme turnover was slowed to minutes within crystals, whereas the reaction was complete within a few milliseconds in solution, and spectra recorded 8 s after mixing were consistent with a ferryl species: either the P_R or the F intermediate. Accordingly, electron density calculated from TR-SFX data measured 8 s after mixing was assigned from the refined Fe-O and Cu_B-O distances as representing the P_R species (Fe^{4+}=O^{2-}; Cu_B^{2+}-OH^-) (*98*) despite P_R normally rising and decaying within microseconds.

Initiating enzymatic reactions in crystals by exchanging reagents within a flow cell dates to pioneering studies of elastase (*1*) and glycogen phosphorylase B (*2*). As the study of O_2 reduction by CcO (*98*) illustrates, serial crystallography in combination with mixing creates a very promising method for recording structural changes in enzymes that are not naturally light sensitive. Other successful mixing experiments include observing electron density changes associated with the binding of inhibitors to lysozyme (*99*, *100*) and the binding of substrate and ring opening catalyzed by xylose isomerase (*100*) and β-lactamase (*97*, *101*). Similar reaction initiation protocols allowed a putative thioimidate intermediate, in which an active site cysteine residue is covalently modified through a nucleophilic attack on an organic isocyanide, to be observed in isocyanide hydratase 15 s after microcrystals were mixed with substrate (*102*). This cysteine modification was also shown to displace a neighboring α helix, and it was speculated that this movement allowed a water molecule to enter into the active site, hydrolyze the thioimidate intermediate, and release the product. In another mixing experiment—in a notable departure from protein-based structural biology—conformational changes within an RNA molecule that regulates mRNA translation through ligand binding (a riboswitch) were reported. TR-SFX data collected 10 s after mixing revealed the rearrangement of active site bases associated with ligand binding (*103*) and larger conformational changes associated with the fully ligated structure were observed after a 600-s delay.

Although it is conceptually straightforward to initiate a TR-SFX or TR-SSX study by mixing microcrystals with appropriate reactants or reagents, and theoretical considerations suggest that the time required for small molecules to diffuse into microcrystals may allow many enzymatic reactions to be rapidly initiated (*104*), this class of experiments frequently encounters many challenges in practice. Most importantly, the reaction must be synchronized, and it is telling that the examples above (*97*, *98*, *100*–*103*) all involved reactions that are slowed to seconds or longer within crystals, in some cases almost stalling at a specific point along the enzymatic reaction pathway. Moreover, for electron density changes to be interpreted relative to enzymatic function, it is helpful if the reaction can be characterized within crystals in advance of x-ray diffraction

studies using ultraviolet-visible spectroscopy (*98*), electron paramagnetic resonance spectroscopy (*105*), or other biophysical methods.

Photocaged substrates overcome one of these limitations, as the caged compound can be infused into crystals in advance of the reaction being initiated by a laser pulse. It has been three decades since this approach first yielded observable electron density changes occurring over several minutes during GTP hydrolysis by Ha-Ras p21 (*3*). Recent TR-SFX data recorded 20 ms after caged nitric oxide (NO) was released by light showed the mode of binding of NO to the active site heme of fungal P450 nitric oxide reductase (*106*). Photocaged fluoroacetate was also used as a source of substrate for the enzyme fluoroacetate dehalogenase (*107*). A sequence of TR-SSX data collected 30 ms to 30 s after the release of fluoroacetate revealed the sequence of substrate binding, the formation of an ester-bonded intermediate, the guided attack of a water molecule that cleaved the ester bond, and the release of the glycolate product (*107*). A novel mechanism of allosteric communication between the active sites of the homodimer also emerged from these TR-SSX studies.

Future trajectories

High-quality difference Fourier electron density maps were first calculated from TR-SFX data 7 years ago (*7*). This milestone sparked a marked acceleration in the rate at which time-dependent structural changes have been reported from macromolecular systems (Fig. 6). TR-SFX also opened up the domain of ultrafast structural biology (Fig. 3), which represents a revolutionary advance given that it was not previously possible to visualize protein conformational changes on the time scale at which chemical bonds break (*59*) or isomerize (*54*, *61*–*64*, *68*, *72*) or at which electrons are transferred between cofactors (*83*). More evolutionary adaptations of mature approaches for reaction initiation (*1*–*3*) to time-resolved serial crystallography include mix-and-inject serial crystallography (*97*–*101*, *103*) and the use of photocages (*106*, *107*). We anticipate that temperature jumps may soon also be used as another tool for reaction initiation (*108*, *109*) in serial crystallography studies. These advances allow the field to expand beyond the constraints of studying light-sensitive proteins and pave the way for elucidating the structural mechanisms of a broad diversity of enzymes.

Synchrotron radiation facilities typically support an order of magnitude more experiments each year than XFELs, and there are many more synchrotron radiation laboratories worldwide. The recent technology transfer of time-resolved serial crystallography to synchrotron radiation facilities (*71*, *99*, *100*, *107*, *110*–*113*), as well as related approaches requiring far fewer crystals (*113*), is therefore critical for the field to develop. A larger user community will also drive standardization that will lower entry barriers for new users. Data processing pipelines could become more streamlined (*33*), and structural modeling and interpretation tools could be developed to strengthen confidence in the structural conclusions drawn from these studies (*83*, *114*, *115*). These advances may ultimately facilitate extraordinary discoveries in life science that cannot be achieved using any other biophysical approach. Some possibilities include allosteric control of chemical pathways (*107*) or a deeper understanding of the role of low-amplitude protein vibrational modes (*21*, *54*, *59*, *68*) in enzyme catalysis. Given the current pace of development, we believe that structural enzymology, both using time-resolved x-ray diffraction and emerging single-particle cryo–electron microscopy approaches (*116*, *117*), will mature into powerful experimental approaches that provide wide-ranging insights into the chemical subtleties of a living cell.

REFERENCES AND NOTES

1. T. Alber, G. A. Petsko, D. Tsernoglou, Crystal structure of elastase–substrate complex at −55 °C. *Nature* **263**, 297–300 (1976). doi: 10.1038/263297a0; pmid: 958484
2. J. Hajdu *et al*., Catalysis in the crystal: Synchrotron radiation studies with glycogen phosphorylase b. *EMBO J.* **6**, 539–546 (1987). doi: 10.1002/j.1460-2075.1987.tb04786.x; pmid: 3107984
3. I. Schlichting *et al*., Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature* **345**, 309–315 (1990). doi: 10.1038/345309a0; pmid: 2111463
4. K. Moffat, Laue diffraction and time-resolved crystallography: A personal history. *Philos. Trans. A Math Phys. Eng. Sci.* **377**, 20180243 (2019). doi: 10.1098/rsta.2018.0243; pmid: 31030647
5. H. N. Chapman *et al*., Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011). doi: 10.1038/nature09750; pmid: 21293373
6. P. Emma *et al*., First lasing and operation of an ångstrom-wavelength free-electron laser. *Nat. Photonics* **4**, 641–647 (2010). doi: 10.1038/nphoton.2010.176
7. J. Tenboer *et al*., Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242–1246 (2014). doi: 10.1126/science.1259357; pmid: 25477465
8. K. Moffat, D. Szebenyi, D. Bilderback, X-ray Laue diffraction from protein crystals. *Science* **223**, 1423–1425 (1984). doi: 10.1126/science.223.4643.1423; pmid: 17746054
9. J. Hajdu *et al*., Millisecond X-ray diffraction and the first electron density map from Laue photographs of a protein crystal. *Nature* **329**, 178–181 (1987). doi: 10.1038/329178a0; pmid: 3114644
10. V. Srajer *et al*., Photolysis of the carbon monoxide complex of myoglobin: Nanosecond time-resolved crystallography. *Science* **274**, 1726–1729 (1996). doi: 10.1126/science.274.5293.1726; pmid: 8939867
11. F. Schotte *et al*., Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science* **300**, 1944–1947 (2003). doi: 10.1126/science.1078797; pmid: 12817148
12. U. K. Genick *et al*., Structure of a protein photocycle intermediate by millisecond time-resolved crystallography. *Science* **275**, 1471–1475 (1997). doi: 10.1126/science.275.5305.1471; pmid: 9045611
13. B. Perman *et al*., Energy transduction on the nanosecond time scale: Early structural events in a xanthopsin photocycle. *Science* **279**, 1946–1950 (1998). doi: 10.1126/science.279.5358.1946; pmid: 9506946
14. F. Schotte *et al*., Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19256–19261 (2012). doi: 10.1073/pnas.1210938109; pmid: 23132943
15. Y. O. Jung *et al*., Volume-conserving *trans*–*cis* isomerization pathways in photoactive yellow protein visualized by picosecond X-ray crystallography. *Nat. Chem.* **5**, 212–220 (2013). doi: 10.1038/nchem.1565; pmid: 23422563
16. J. E. Knapp, R. Pahl, V. Srajer, W. E. Royer Jr., Allosteric action in real time: Time-resolved crystallographic studies of a cooperative dimeric hemoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7649–7654 (2006). doi: 10.1073/pnas.0509411103; pmid: 16684887
17. F. Schotte *et al*., Real-time tracking of CO migration and binding in the α and β subunits of human hemoglobin via 150-ps time-resolved Laue crystallography. *Chem. Phys.* **422**, 98–106 (2013). doi: 10.1016/j.chemphys.2012.12.030; pmid: 24839343
18. J. Key, V. Srajer, R. Pahl, K. Moffat, Time-resolved crystallographic studies of the heme domain of the oxygen sensor FixL: Structural dynamics of ligand rebinding and their relation to signal transduction. *Biochemistry* **46**, 4706–4715 (2007). doi: 10.1021/bi700043c; pmid: 17385895
19. A. B. Wöhri *et al*., Light-induced structural changes in a photosynthetic reaction center caught by Laue diffraction. *Science* **328**, 630–633 (2010). doi: 10.1126/science.1186159; pmid: 20431017
20. R. H. Baxter *et al*., Time-resolved crystallographic studies of light-induced structural changes in the photosynthetic reaction center. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5982–5987 (2004). doi: 10.1073/pnas.0306840101; pmid: 15073325
21. D. R. Hekstra *et al*., Electric-field-stimulated protein mechanics. *Nature* **540**, 400–405 (2016). doi: 10.1038/nature20571; pmid: 27926732
22. M. Cammarata *et al*., Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering. *Nat. Methods* **5**, 881–886 (2008). doi: 10.1038/nmeth.1255; pmid: 18806790
23. M. Andersson *et al*., Structural dynamics of light-driven proton pumps. *Structure* **17**, 1265–1275 (2009). doi: 10.1016/j.str.2009.07.007; pmid: 19748347
24. H. S. Cho *et al*., Protein structural dynamics in solution unveiled via 100-ps time-resolved x-ray scattering. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 7281–7286 (2010). doi: 10.1073/pnas.1002951107; pmid: 20406909
25. T. W. Kim *et al*., Combined probes of X-ray scattering and optical spectroscopy reveal how global conformational change is temporally and spatially linked to local structural perturbation in photoactive yellow protein. *Phys. Chem. Chem. Phys.* **18**, 8911–8919 (2016). doi: 10.1039/C6CP00476H; pmid: 26960811
26. A. Björling *et al*., Structural photoactivation of a full-length bacterial phytochrome. *Sci. Adv.* **2**, e1600920 (2016). doi: 10.1126/sciadv.1600920; pmid: 27536728
27. H. E. Huxley, K. C. Holmes, Development of synchrotron radiation as a high-intensity source for X-ray diffraction. *J. Synchrotron Radiat.* **4**, 366–379 (1997). doi: 10.1107/S0909049597011618; pmid: 16699251
28. R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000). doi: 10.1038/35021099; pmid: 10963603
29. H. N. Chapman, C. Caleman, N. Timneanu, Diffraction before destruction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130313 (2014). doi: 10.1098/rstb.2013.0313; pmid: 24914146
30. A. Barty *et al*., Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat. Photonics* **6**, 35–40 (2012). doi: 10.1038/nphoton.2011.297; pmid: 24078834
31. D. P. DePonte *et al*., Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D Appl. Phys.* **41**, 195505 (2008). doi: 10.1088/0022-3727/41/19/195505
32. U. Weierstall, J. C. Spence, R. B. Doak, Injector for scattering measurements on fully solvated biospecies. *Rev. Sci. Instrum.* **83**, 035108 (2012). doi: 10.1063/1.3693040; pmid: 22462961
33. T. White *et al*., *CrystFEL*: A software suite for snapshot serial crystallography. *J. Appl. Cryst.* **45**, 335–341 (2012). doi: 10.1107/S0021889812002312
34. L. Lomb *et al*., Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Phys. Rev. B* **84**, 214111 (2011). doi: 10.1103/PhysRevB.84.214111; pmid: 24089594

35. L. C. Johansson *et al*., Lipidic phase membrane protein serial femtosecond crystallography. *Nat. Methods* **9**, 263–265 (2012). doi: 10.1038/nmeth.1867; pmid: 22286383
36. S. Boutet *et al*., High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364 (2012). doi: 10.1126/science.1217737; pmid: 22653729
37. L. Redecke *et al*., Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an x-ray laser. *Science* **339**, 227–230 (2013). doi: 10.1126/science.1229663; pmid: 23196907
38. W. Liu *et al*., Serial femtosecond crystallography of G protein–coupled receptors. *Science* **342**, 1521–1524 (2013). doi: 10.1126/science.1244142; pmid: 24357322
39. S. Boutet, G. J. Williams, The Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source (LCLS). *New J. Phys.* **12**, 035024 (2010). doi: 10.1088/1367-2630/12/3/035024
40. I. Schlichting, Serial femtosecond crystallography: The first five years. *IUCrJ* **2**, 246–255 (2015). doi: 10.1107/S205225251402702X; pmid: 25866661
41. H. N. Chapman, Structure determination using X-ray free-electron laser pulses. *Methods Mol. Biol.* **1607**, 295–324 (2017). doi: 10.1007/978-1-4939-7000-1_12; pmid: 28573578
42. L. C. Johansson, B. Stauch, A. Ishchenko, V. Cherezov, A bright future for serial femtosecond crystallography with XFELs. *Trends Biochem. Sci.* **42**, 749–762 (2017). doi: 10.1016/j.tibs.2017.06.007; pmid: 28733116
43. T. R. Barends *et al*., *De novo* protein crystal structure determination from X-ray free-electron laser data. *Nature* **505**, 244–247 (2014). doi: 10.1038/nature12773; pmid: 24270807
44. K. Nass *et al*., Protein structure determination by single-wavelength anomalous diffraction phasing of X-ray free-electron laser data. *IUCrJ* **3**, 180–191 (2016). doi: 10.1107/S2052252516002980; pmid: 27158504
45. J. P. Colletier *et al*., *De novo* phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature* **539**, 43–47 (2016). doi: 10.1038/nature19825; pmid: 27680699
46. I. Schlichting, Experimental phasing of serial femtosecond crystallography data. *IUCrJ* **4**, 516–517 (2017). doi: 10.1107/S2052252517012167; pmid: 28989708
47. K. Ayyer *et al*., Macromolecular diffractive imaging using imperfect crystals. *Nature* **530**, 202–206 (2016). doi: 10.1038/nature16949; pmid: 26863980
48. A. J. Morgan *et al*., *Ab initio* phasing of the diffraction of crystals with translational disorder. *Acta Crystallogr. A* **75**, 25–40 (2019). doi: 10.1107/S2053273318015395; pmid: 30575581
49. J. J. van Thor, Advances and opportunities in ultrafast X-ray crystallography and ultrafast structural optical crystallography of nuclear and electronic protein dynamics. *Struct. Dyn.* **6**, 050901 (2019). doi: 10.1063/1.5110685; pmid: 31559317
50. R. Henderson, J. K. Moffat, The difference Fourier technique in protein crystallography: Errors and their treatment. *Acta Crystallogr. B* **27**, 1414–1420 (1971). doi: 10.1107/S0567740871004060
51. M. Weik, J. P. Colletier, Temperature-dependent macromolecular X-ray crystallography. *Acta Crystallogr. D* **66**, 437–446 (2010). doi: 10.1107/S0907444910002702; pmid: 20382997
52. S. Gozem, H. L. Luk, I. Schapiro, M. Olivucci, Theory and simulation of the ultrafast double-bond isomerization of biological chromophores. *Chem. Rev.* **117**, 13502–13565 (2017). doi: 10.1021/acs.chemrev.7b00177; pmid: 29083892
53. R. J. D. Miller, O. Paré-Labrosse, A. Sarracini, J. E. Besaw, Three-dimensional view of ultrafast dynamics in photoexcited bacteriorhodopsin in the multiphoton regime and biological relevance. *Nat. Commun.* **11**, 1240 (2020). doi: 10.1038/s41467-020-14971-0; pmid: 32144255
54. G. Nass Kovacs *et al*., Three-dimensional view of ultrafast dynamics in photoexcited bacteriorhodopsin. *Nat. Commun.* **10**, 3177 (2019). doi: 10.1038/s41467-019-10758-0; pmid: 31320619
55. C. D. M. Hutchison *et al*., Photocycle populations with femtosecond excitation of crystalline photoactive yellow protein. *Chem. Phys. Lett.* **654**, 63–71 (2016). doi: 10.1016/j.cplett.2016.04.087
56. D. Arnlund *et al*., Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nat. Methods* **11**, 923–926 (2014). doi: 10.1038/nmeth.3067; pmid: 25108686
57. M. Levantino *et al*., Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. *Nat. Commun.* **6**, 6772 (2015). doi: 10.1038/ncomms7772; pmid: 25832715
58. L. U. Brinkmann, J. S. Hub, Ultrafast anisotropic protein quake propagation after CO photodissociation in myoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10565–10570 (2016). doi: 10.1073/pnas.1603539113; pmid: 27601659
59. T. R. Barends *et al*., Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350**, 445–450 (2015). doi: 10.1126/science.aac5492; pmid: 26359336
60. M. Levantino *et al*., Observing heme doming in myoglobin with femtosecond X-ray absorption spectroscopy. *Struct. Dyn.* **2**, 041713 (2015). doi: 10.1063/1.4921907; pmid: 26798812
61. K. Pande *et al*., Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725–729 (2016). doi: 10.1126/science.aad5081; pmid: 27151871
62. S. Pandey *et al*., Time-resolved serial femtosecond crystallography at the European XFEL. *Nat. Methods* **17**, 73–78 (2020). doi: 10.1038/s41592-019-0628-z; pmid: 31740816
63. N. Coquelle *et al*., Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* **10**, 31–37 (2018). doi: 10.1038/nchem.2853; pmid: 29256511
64. E. Claesson *et al*., The primary structural photoresponse of phytochrome proteins captured by a femtosecond X-ray laser. *eLife* **9**, e53514 (2020). doi: 10.7554/eLife.53514; pmid: 32228856
65. J. Woodhouse *et al*., Photoswitching mechanism of a fluorescent protein revealed by time-resolved crystallography and transient absorption spectroscopy. *Nat. Commun.* **11**, 741 (2020). doi: 10.1038/s41467-020-14537-0; pmid: 32029745
66. J. Hughes, Rapid response. *eLife* **9**, e57105 (2020). doi: 10.7554/eLife.57105; pmid: 32301437
67. M. Carrillo *et al*., High-resolution crystal structures of transient intermediates in the phytochrome photocycle. *Structure* **29**, 743–754.e4 (2021). doi: 10.1016/j.str.2021.03.004; pmid: 33756101
68. P. Nogly *et al*., Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* **361**, eaat0094 (2018). doi: 10.1126/science.aat0094; pmid: 29903883
69. C. Wickstrand *et al*., A tool for visualizing protein motions in time-resolved crystallography. *Struct. Dyn.* **7**, 024701 (2020). doi: 10.1063/1.5126921; pmid: 32266303
70. E. Nango *et al*., A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* **354**, 1552–1557 (2016). doi: 10.1126/science.aah3497; pmid: 28008064
71. T. Weinert *et al*., Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography. *Science* **365**, 61–65 (2019). doi: 10.1126/science.aaw8634; pmid: 31273117
72. P. Skopintsev *et al*., Femtosecond-to-millisecond structural changes in a light-driven sodium pump. *Nature* **583**, 314–318 (2020). doi: 10.1038/s41586-020-2307-8; pmid: 32499654
73. J. H. Yun *et al*., Early-stage dynamics of chloride ion-pumping rhodopsin revealed by a femtosecond X-ray laser. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020486118 (2021). doi: 10.1073/pnas.2020486118; pmid: 33753488
74. K. Oda *et al*., Time-resolved serial femtosecond crystallography reveals early structural changes in channelrhodopsin. *eLife* **10**, e62389 (2021). doi: 10.7554/eLife.62389; pmid: 33752801
75. J. Kern *et al*., Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. *Science* **340**, 491–495 (2013). doi: 10.1126/science.1234273; pmid: 23413188
76. J. Kern *et al*., Taking snapshots of photosynthetic water oxidation using femtosecond X-ray diffraction and spectroscopy. *Nat. Commun.* **5**, 4371 (2014). doi: 10.1038/ncomms5371; pmid: 25006873
77. C. Kupitz *et al*., Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513**, 261–265 (2014). doi: 10.1038/nature13453; pmid: 25043005
78. I. D. Young *et al*., Structure of photosystem II and substrate binding at room temperature. *Nature* **540**, 453–457 (2016). doi: 10.1038/nature20161; pmid: 27871088
79. M. Suga *et al*., Light-induced structural changes and the site of O=O bond formation in PSII caught by XFEL. *Nature* **543**, 131–135 (2017). doi: 10.1038/nature21400; pmid: 28219079
80. J. Kern *et al*., Structures of the intermediates of Kok's photosynthetic water oxidation clock. *Nature* **563**, 421–425 (2018). doi: 10.1038/s41586-018-0681-2; pmid: 30405241
81. M. Suga *et al*., An oxyl/oxo mechanism for oxygen-oxygen coupling in PSII revealed by an x-ray free-electron laser. *Science* **366**, 334–338 (2019). doi: 10.1126/science.aax6998; pmid: 31624207
82. M. Ibrahim *et al*., Untangling the sequence of events during the $S_2 \rightarrow S_3$ transition in photosystem II and implications for the water oxidation mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12624–12635 (2020). doi: 10.1073/pnas.2000529117; pmid: 32434915
83. R. Dods *et al*., Ultrafast structural changes within a photosynthetic reaction centre. *Nature* **589**, 310–314 (2021). doi: 10.1038/s41586-020-3000-7; pmid: 33268896
84. Y. Umena, K. Kawakami, J. R. Shen, N. Kamiya, Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **473**, 55–60 (2011). doi: 10.1038/nature09913; pmid: 21499260
85. B. Kok, B. Forbush, M. McGloin, Cooperation of charges in photosynthetic O_2 evolution-I. A linear four step mechanism. *Photochem. Photobiol.* **11**, 457–475 (1970). doi: 10.1111/j.1751-1097.1970.tb06017.x; pmid: 5456273
86. N. K. Sauter *et al*., No observable conformational changes in PSII. *Nature* **533**, E1–E2 (2016). doi: 10.1038/nature17983; pmid: 27193689
87. F. D. Fuller *et al*., Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat. Methods* **14**, 443–449 (2017). doi: 10.1038/nmeth.4195; pmid: 28250468
88. K. Nass *et al*., Structural dynamics in proteins induced by and probed with X-ray free-electron laser pulses. *Nat. Commun.* **11**, 1814 (2020). doi: 10.1038/s41467-020-15610-4; pmid: 32286284
89. S. C. Jensen *et al*., X-ray emission spectroscopy at X-ray free electron lasers: Limits to observation of the classical spectroscopic response for electronic structure analysis. *J. Phys. Chem. Lett.* **10**, 441–446 (2019). doi: 10.1021/acs.jpclett.8b03595; pmid: 30566358
90. M. Suga *et al*., Native structure of photosystem II at 1.95 Å resolution viewed by femtosecond X-ray pulses. *Nature* **517**, 99–103 (2015). doi: 10.1038/nature13991; pmid: 25470056
91. V. Krewald *et al*., Metal oxidation states in biological water splitting. *Chem. Sci.* **6**, 1676–1695 (2015). doi: 10.1039/C4SC03720K; pmid: 29308133
92. P. E. Siegbahn, Water oxidation mechanism in photosystem II, including oxidations, proton release pathways, O-O bond formation and O_2 release. *Biochim. Biophys. Acta* **1827**, 1003–1019 (2013). doi: 10.1016/j.bbabio.2012.10.006; pmid: 23103385
93. K. Hirata *et al*., Determination of damage-free crystal structure of an X-ray-sensitive protein using an XFEL. *Nat. Methods* **11**, 734–736 (2014). doi: 10.1038/nmeth.2962; pmid: 24813624
94. I. Ishigami *et al*., Crystal structure of CO-bound cytochrome *c* oxidase determined by serial femtosecond X-ray crystallography at room temperature. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8011–8016 (2017). doi: 10.1073/pnas.1705628114; pmid: 28698372
95. R. Andersson *et al*., Serial femtosecond crystallography structure of cytochrome c oxidase at room temperature. *Sci. Rep.* **7**, 4518 (2017). doi: 10.1038/s41598-017-04817-z; pmid: 28674417
96. A. Shimada *et al*., A nanosecond time-resolved XFEL analysis of structural changes associated with CO release from cytochrome c oxidase. *Sci. Adv.* **3**, e1603042 (2017). doi: 10.1126/sciadv.1603042; pmid: 28740863
97. C. Kupitz *et al*., Structural enzymology using X-ray free electron lasers. *Struct. Dyn.* **4**, 044003 (2016). doi: 10.1063/1.4972069; pmid: 28083542
98. I. Ishigami *et al*., Snapshot of an oxygen intermediate in the catalytic reaction of cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 3572–3577 (2019). doi: 10.1073/pnas.1814526116; pmid: 30808749
99. K. R. Beyerlein *et al*., Mix-and-diffuse serial synchrotron crystallography. *IUCrJ* **4**, 769–777 (2017). doi: 10.1107/S2052252517013124; pmid: 29123679
100. P. Mehrabi *et al*., Liquid application method for time-resolved analyses by serial synchrotron crystallography. *Nat. Methods* **16**, 979–982 (2019). doi: 10.1038/s41592-019-0553-1; pmid: 31527838
101. J. L. Olmos Jr. *et al*., Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography. *BMC Biol.* **16**, 59 (2018). doi: 10.1186/s12915-018-0524-5; pmid: 29848358
102. M. Dasgupta *et al*., Mix-and-inject XFEL crystallography reveals gated conformational dynamics during enzyme

catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25634–25640 (2019). doi: 10.1073/pnas.1901864116; pmid: 31801874
103. J. R. Stagno *et al*., Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541**, 242–246 (2017). doi: 10.1038/nature20599; pmid: 27841871
104. M. Schmidt, Mix and inject: Reaction initiation by diffusion for time-resolved macromolecular crystallography. *Adv. Cond. Mat. Phys.* **2013**, 167276 (2013). doi: 10.1155/2013/167276
105. G. D. Calvey, A. M. Katz, K. A. Zielinski, B. Dzikovski, L. Pollack, Characterizing enzyme reactions in microcrystals for effective mix-and-inject experiments using x-ray free-electron lasers. *Anal. Chem.* **92**, 13864–13870 (2020). doi: 10.1021/acs.analchem.0c02569; pmid: 32955854
106. T. Tosha *et al*., Capturing an initial intermediate during the P450nor enzymatic reaction using time-resolved XFEL crystallography and caged-substrate. *Nat. Commun.* **8**, 1585 (2017). doi: 10.1038/s41467-017-01702-1; pmid: 29147002
107. P. Mehrabi *et al*., Time-resolved crystallography reveals allosteric communication aligned with molecular breathing. *Science* **365**, 1167–1170 (2019). doi: 10.1126/science.aaw9904; pmid: 31515393
108. D. A. Keedy *et al*., Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife* **4**, e07574 (2015). doi: 10.7554/eLife.07574; pmid: 26422513
109. M. C. Thompson *et al*., Temperature-jump solution X-ray scattering reveals distinct motions in a dynamic enzyme. *Nat. Chem.* **11**, 1058–1066 (2019). doi: 10.1038/s41557-019-0329-3; pmid: 31527847
110. P. Mehrabi *et al*., The HARE chip for efficient time-resolved serial synchrotron crystallography. *J. Synchrotron Radiat.* **27**, 360–370 (2020). doi: 10.1107/S1600577520000685; pmid: 32153274
111. D. C. F. Monteiro *et al*., 3D-MiXD: 3D-printed X-ray-compatible microfluidic devices for rapid, low-consumption serial synchrotron crystallography data collection in flow. *IUCrJ* **7**, 207–219 (2020). doi: 10.1107/S2052252519016865; pmid: 32148849
112. E. C. Schulz *et al*., The hit-and-return system enables efficient time-resolved serial synchrotron crystallography. *Nat. Methods* **15**, 901–904 (2018). doi: 10.1038/s41592-018-0180-2; pmid: 30377366
113. S. Aumonier *et al*., Millisecond time-resolved serial oscillation crystallography of a blue-light photoreceptor at a synchrotron. *IUCrJ* **7**, 728–736 (2020). doi: 10.1107/S2052252520007411; pmid: 32695419
114. T. Ursby, D. Bourgeois, Improved estimation of structure-factor difference amplitudes from poorly accurate data. *Acta Crystallogr. A* **53**, 564–575 (1997). doi: 10.1107/S0108767397004522
115. S. Rajagopal, M. Schmidt, S. Anderson, H. Ihee, K. Moffat, Analysis of experimental time-resolved crystallographic data by singular value decomposition. *Acta Crystallogr. D* **60**, 860–871 (2004). doi: 10.1107/S0907444904004160; pmid: 15103131
116. J. Frank, A. Ourmazd, Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods* **100**, 61–67 (2016). doi: 10.1016/j.ymeth.2016.02.007; pmid: 26884261
117. J. Frank, Time-resolved cryo-electron microscopy: Recent progress. *J. Struct. Biol.* **200**, 303–306 (2017). doi: 10.1016/j.jsb.2017.06.005; pmid: 28625887
118. U. Weierstall *et al*., Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 3309 (2014). doi: 10.1038/ncomms4309; pmid: 24525480
119. D. Wang, U. Weierstall, L. Pollack, J. Spence, Double-focusing mixing jet for XFEL study of chemical kinetics. *J. Synchrotron Radiat.* **21**, 1364–1366 (2014). doi: 10.1107/S160057751401858X; pmid: 25343806
120. G. D. Calvey, A. M. Katz, C. B. Schaffer, L. Pollack, Mixing injector enables time-resolved crystallography with high hit rate at X-ray free electron lasers. *Struct. Dyn.* **3**, 054301 (2016). doi: 10.1063/1.4961971; pmid: 27679802
121. D. Oberthuer *et al*., Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Sci. Rep.* **7**, 44628 (2017). doi: 10.1038/srep44628; pmid: 28300169
122. J. Knoška *et al*., Ultracompact 3D microfluidics for time-resolved structural biology. *Nat. Commun.* **11**, 657 (2020). doi: 10.1038/s41467-020-14434-6; pmid: 32005876
123. D. A. Sherrell *et al*., A modular and compact portable mini-endstation for high-precision, high-speed fixed target serial crystallography at FEL and synchrotron sources. *J. Synchrotron Radiat.* **22**, 1372–1378 (2015). doi: 10.1107/S1600577515016938; pmid: 26524301
124. C. Mueller *et al*., Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography. *Struct. Dyn.* **2**, 054302 (2015). doi: 10.1063/1.4928706; pmid: 26798825
125. U. K. Genick, Structure-factor extrapolation using the scalar approximation: Theory, applications and limitations. *Acta Crystallogr. D* **63**, 1029–1041 (2007). doi: 10.1107/S0907444907038164; pmid: 17881820
126. M. L. Grünbein *et al*., Effect of X-ray free-electron laser-induced shockwaves on haemoglobin microcrystals delivered in a liquid jet. *Nat. Commun.* **12**, 1672 (2021). doi: 10.1038/s41467-021-21819-8; pmid: 33723266
127. C. N. Lincoln, A. E. Fitzpatrick, J. J. van Thor, Photoisomerisation quantum yield and non-linear cross-sections with femtosecond excitation of the photoactive yellow protein. *Phys. Chem. Chem. Phys.* **14**, 15752–15764 (2012). doi: 10.1039/c2cp41718a; pmid: 23090503
128. V. I. Prokhorenko, A. Halpin, P. J. Johnson, R. J. Miller, L. S. Brown, Coherent control of the isomerization of retinal in bacteriorhodopsin in the high intensity regime. *J. Chem. Phys.* **134**, 085105 (2011). doi: 10.1063/1.3554743; pmid: 21361560
129. M. L. Grünbein *et al*., Illumination guidelines for ultrafast pump-probe experiments by serial femtosecond crystallography. *Nat. Methods* **17**, 681–684 (2020). doi: 10.1038/s41592-020-0847-3; pmid: 32451477
130. C. Gati *et al*., Serial crystallography on in vivo grown microcrystals using synchrotron radiation. *IUCrJ* **1**, 87–94 (2014). doi: 10.1107/S2052252513033939; pmid: 25075324

ACKNOWLEDGMENTS

We thank P. Nogly, M. Schmidt, J. Standfuss, and T. Weinert for providing difference electron density maps and M. Cammarata, A. Björling, H. Ihee, and S. Westenhoff for providing experimental data or estimates of the deposited heat recorded in TR-XSS studies. **Funding:** This work was supported by the European Union's Horizon 2020 research and innovation program (grant 789030), the Swedish Research Council (grants 2015-00560 and 2017-06734), and the Swedish Foundation for Strategic Research (grant ID17-0060). **Competing interests:** None declared.

10.1126/science.aba0954



Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**

Science Advances
AAAS

GOLD OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

REVIEW SUMMARY

CORONAVIRUS

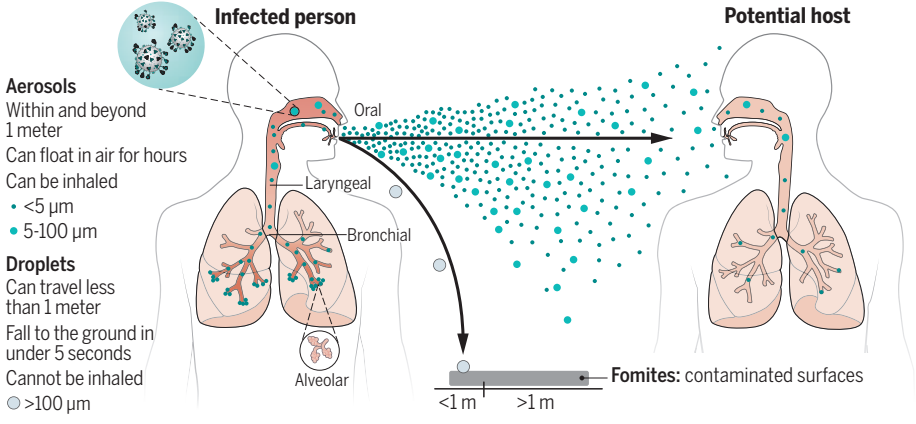
Airborne transmission of respiratory viruses

Chia C. Wang*, Kimberly A. Prather*, Josué Sznitman, Jose L. Jimenez, Seema S. Lakdawala, Zeynep Tufekci, Linsey C. Marr

BACKGROUND: Exposure to droplets produced in the coughs and sneezes of infected individuals or contact with droplet-contaminated surfaces (fomites) have been widely perceived as the dominant transmission modes for respiratory pathogens. Airborne transmission is traditionally defined as involving the inhalation of infectious aerosols or "droplet nuclei" smaller than 5 μm and mainly at a distance of >1 to 2 m away from the infected individual, and such transmission has been thought to be relevant only for "unusual" diseases. However, there is robust evidence supporting the airborne transmission of many respiratory viruses, including severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome (MERS)–CoV, influenza virus, human rhinovirus, and respiratory syncytial virus (RSV). The limitations of traditional views of droplet, fomite, and airborne transmission were illuminated during the COVID-19 pandemic. Droplet and fomite transmission of SARS-CoV-2 alone cannot account for the numerous superspreading events and differences in transmission between indoor and outdoor environments observed during the COVID-19 pandemic. Controversy surrounding how COVID-19 is transmitted and what interventions are needed to control the pandemic has revealed a critical need to better understand the airborne transmission pathway of respiratory viruses, which will allow for better-informed strategies to mitigate the transmission of respiratory infections.

ADVANCES: Respiratory droplets and aerosols can be generated by various expiratory activities. Advances in aerosol measurement techniques, such as aerodynamic and scanning mobility particle sizing, have shown that the majority of exhaled aerosols are smaller than 5 μm, and a large fraction are <1 μm for most respiratory activities, including those produced during breathing, talking, and coughing. Exhaled aerosols occur in multiple size modes that are associated with different generation sites and production mechanisms in the respiratory tract. Although 5 μm has been used historically to distinguish aerosols from droplets, the size distinction between aerosols and droplets should be 100 μm, which represents the largest particle size that can remain suspended in still air for more than 5 s from a height of 1.5 m, typically reach a distance of 1 to 2 m from the emitter (depending on the velocity of airflow carrying the aerosols), and can be inhaled. Aerosols produced by an infected individual may contain infectious viruses, and studies have shown that viruses are enriched in small aerosols (<5 μm). The transport of virus-laden aerosols is affected by the physicochemical properties of aerosols themselves and environmental factors, including temperature, relative humidity, ultraviolet radiation, airflow, and ventilation. Once inhaled, virus-laden aerosols can deposit in different parts of the respiratory tract. Larger aerosols tend to be deposited in the upper airway; however, smaller aerosols, although they can also be deposited there, can penetrate deep into the alveolar region of the lungs. The strong effect of ventilation on transmission, the distinct difference between indoor and outdoor transmission, well-documented long-range transmission, the observed transmission of SARS-CoV-2 despite the use of masks and eye protection, the high frequency of indoor superspreading events of SARS-CoV-2, animal experiments, and airflow simulations provide strong and unequivocal evidence for airborne transmission. Fomite transmission of SARS-CoV-2 has been found to be far less efficient, and droplets are only dominant when individuals are within 0.2 m of each other when talking. Although both aerosols and droplets can be produced by infected individuals during expiratory activities, droplets fall quickly to the ground or surfaces within seconds, leaving an enrichment of aerosols over droplets. The airborne pathway likely contributes to the spread of other respiratory viruses whose transmission was previously characterized as droplet driven. The World Health Organization (WHO) and the US Centers for Disease Control and Prevention (CDC) have officially acknowledged the inhalation of virus-laden aerosols as a main transmission mode in spreading COVID-19 at both short and long ranges in 2021.

OUTLOOK: Airborne transmission of pathogens has been vastly underappreciated, mostly because of an insufficient understanding about the airborne behavior of aerosols and at least partially because of the misattribution of anecdotal observations. Given the lack of evidence for droplet and fomite transmission and the increasingly strong evidence for aerosols in transmitting numerous respiratory viruses, we must acknowledge that airborne transmission is much more prevalent than previously recognized. Given all that we have learned about SARS-CoV-2 infection, the aerosol transmission pathway needs to be reevaluated for all respiratory infectious diseases. Additional precautionary measures must be implemented for mitigating aerosol transmission at both short and long ranges, with particular attention to ventilation, airflows, air filtration, UV disinfection, and mask fit. These interventions are critical tools for ending the current pandemic and preventing future outbreaks. ■



CREDIT: N. CARY/*SCIENCE*

Phases involved in airborne transmission of respiratory viruses. Virus-laden aerosols (<100 μm) are first generated by an infected individual through expiratory activities, through which they are exhaled and transported in the environment. They may be inhaled by a potential host to initiate a new infection, provided that they remain infectious. In contrast to droplets (>100 μm), aerosols can linger in air for hours and travel beyond 1 to 2 m from the infected individual who exhales them, causing new infections at both short and long ranges.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: chiawang@mail.nsysu.edu.tw (C.C.W.); kprather@ucsd.edu (K.A.P.)
This is an open-access article distributed under the terms of the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Cite this article as C. C. Wang *et al.*, *Science* **373**, eabd9149 (2021). DOI: 10.1126/science.abd9149

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd9149

REVIEW

CORONAVIRUS

Airborne transmission of respiratory viruses

Chia C. Wang[1,2]*, Kimberly A. Prather[3]*, Josué Sznitman[4], Jose L. Jimenez[5], Seema S. Lakdawala[6], Zeynep Tufekci[7], Linsey C. Marr[8]

The COVID-19 pandemic has revealed critical knowledge gaps in our understanding of and a need to update the traditional view of transmission pathways for respiratory viruses. The long-standing definitions of droplet and airborne transmission do not account for the mechanisms by which virus-laden respiratory droplets and aerosols travel through the air and lead to infection. In this Review, we discuss current evidence regarding the transmission of respiratory viruses by aerosols—how they are generated, transported, and deposited, as well as the factors affecting the relative contributions of droplet-spray deposition versus aerosol inhalation as modes of transmission. Improved understanding of aerosol transmission brought about by studies of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection requires a reevaluation of the major transmission pathways for other respiratory viruses, which will allow better-informed controls to reduce airborne transmission.

Over the past century, respiratory viruses were thought to be spread mainly through large respiratory droplets, produced in the coughs and sneezes of infected individuals that deposit on the mucous membranes of the eyes, nose, or mouth of potential hosts (droplet transmission) or that deposit on surfaces that are then touched by potential hosts and transferred to mucous membranes (fomite transmission). Such droplets are thought to fall to the ground within 1 to 2 m of the infectious person—a key assumption used by most public health agencies in recommending a safe distance from people infected with respiratory viruses. Thought to be less common, airborne transmission refers to the inhalation of infectious aerosols or "droplet nuclei" (droplets that evaporate in the air), often defined to be smaller than 5 μm and traveling distances of >1 to 2 m away from the infected individual. Aerosols are microscopic liquid, solid, or semisolid particles that are so small that they remain suspended in air. Respiratory aerosols are produced during all expiratory activities, including breathing, talking, singing, shouting, coughing, and sneezing from both healthy individuals and those with respiratory infections (*1*–*4*).

[1]Aerosol Science Research Center, National Sun Yat-sen University, Kaohsiung, Taiwan 804, Republic of China. [2]Department of Chemistry, National Sun Yat-sen University, Kaohsiung, Taiwan 804, Republic of China. [3]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92037, USA. [4]Department of Biomedical Engineering, Israel Institute of Technology, Haifa 32000, Israel. [5]Department of Chemistry and CIRES, University of Colorado, Boulder, CO 80309, USA. [6]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA 15219, USA. [7]School of Information and Department of Sociology, University of North Carolina, Chapel Hill, NC 27599, USA. [8]Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, USA.
***Corresponding author. Email: chiawang@mail.nsysu.edu.tw (C.C.W.); kprather@ucsd.edu (K.A.P.)**

The historical definition of airborne transmission ignores the possibility that aerosols can also be inhaled at close range to an infected person, where exposure is more likely because exhaled aerosols are more concentrated closer to the person emitting them. Moreover, rather than the conventional definition of 5 μm, it has recently been suggested that the size distinction between aerosols and droplets should be updated to 100 μm, as this distinguishes between the two on the basis of their aerodynamic behavior (*5*–*7*). Specifically, 100 μm represents the largest particles that remain suspended in still air for >5 s (from a height of 1.5 m), travel beyond 1 m from the infectious person, and can be inhaled. Although droplets produced by an infectious individual through coughing or sneezing may convey infection at short distances (<0.5 m), the number and viral load of aerosols produced through speaking and other expiratory activities are much higher than those of droplets (*8*–*10*). Aerosols are small enough to linger in air, accumulate in poorly ventilated spaces, and be inhaled at both short and long ranges, calling for an urgent need to include aerosol precautions in current respiratory disease control protocols. During the COVID-19 pandemic, controls have focused mainly on protecting against droplet and fomite transmission, whereas the airborne route has required much more evidence before controls can be added to protect against it.

Debates surrounding the relative importance of different transmission modes in spreading respiratory disease have spanned centuries. Before the 20th century, infectious respiratory diseases were thought to spread by "pestilential particles" released by infected individuals (*11*, *12*). This view of airborne transmission was dismissed in the early 1900s by Charles Chapin, who claimed that contact was the chief route for respiratory disease transmission, with spray-borne (droplet) transmission being an extension of contact transmission (*13*). Chapin was concerned that mentioning transmission by air would scare people into inaction and displace hygiene practices. Chapin erroneously equated infections at close range with droplet transmission—neglecting the fact that aerosol transmission also occurs at short distances. This unsupported assumption became widespread in epidemiological studies (*14*), and mitigation strategies for controlling respiratory virus transmission have since focused on limiting droplet and fomite transmission (*15*). Some of these strategies are also partially effective for limiting aerosol transmission, leading to the erroneous conclusion that their efficacy proved droplet transmission.

Despite the assumed dominance of droplet transmission, there is robust evidence supporting the airborne transmission of many respiratory viruses, including measles virus (*16*–*18*), influenza virus (*19*–*24*), respiratory syncytial virus (RSV) (*25*), human rhinovirus (hRV) (*9*, *26*–*28*), adenovirus, enterovirus (*29*), severe acute respiratory syndrome coronavirus (SARS-CoV) (*30*, *31*), Middle East respiratory syndrome coronavirus (MERS-CoV) (*32*), and SARS-CoV-2 (*33*–*36*) (Table 1). Airborne transmission has been estimated to account for approximately half of the transmission of influenza A virus in one study of a household setting (*20*). A human challenge study on rhinovirus transmission concluded that aerosols were likely the dominant transmission mode (*26*). SARS-CoV-2 infection of hamsters and ferrets has been shown to transmit through air in experimental configurations designed to exclude contributions from direct contact and droplet transmission (*33*, *37*, *38*). Analysis of respiratory emissions during infection with influenza virus, parainfluenza virus, RSV, human metapneumovirus, and hRV has revealed the presence of viral genomes in a variety of aerosol sizes, with the highest amount detected in aerosols <5 μm rather than in larger aerosols (*39*). SARS-CoV-2 RNA has been detected and infectious virus has been recovered in aerosols ranging from 0.25 to >4 μm (*34*, *35*, *40*–*44*). Influenza virus RNA has also been detected in both fine (≤5 μm) and coarse (>5 μm) aerosols exhaled from infected individuals, with more viral RNA contained in the fine aerosol particles (*23*). Laboratory studies have found that aerosolized SARS-CoV-2 has a half-life of ~1 to 3 hours (*45*–*47*). The World Health Organization (WHO) and the US Centers for Disease Control and Prevention (CDC) officially acknowledged inhalation of virus-laden aerosols as a main mode in spreading SARS-CoV-2 at both short and long ranges in April and May of 2021, respectively (*48*, *49*).

Mathematical modeling of exposure to respiratory pathogens supports that transmission

Table 1. Airborne transmission of respiratory viruses. Representative evidence of airborne transmission for various respiratory viruses and their basic reproduction number. Cells with dashes indicate not applicable.

Virus name	Scope of studies and/or approaches							Basic reproduction number (R_0)
	Air sampling and PCR	Air sampling and cell culture	Animal models	Laboratory or clinical studies	Epidemiological analysis	Simulation and modeling	Size-resolved information	
SARS-CoV	(*31*)	(*31*)	–	(*30*)	(*30*)	(*30*)	–	2.0–3.0 (*197*)
MERS-CoV	(*32*)	(*32*, *103*)	(*103*, *198*)	(*32*)	–	–	–	0.50–0.92 (*197*)
SARS-CoV-2	(*41*–*44*)	(*34*, *35*, *40*)	(*33*, *37*, *199*)	(*34*, *45*, *107*)	(*36*, *64*, *71*, *72*, *186*)	(*36*, *50*)	(*34*, *41*, *43*)	1.4–8.9 (*57*, *58*)
Influenza virus	(*22*, *23*, *98*, *102*, *106*)	(*23*, *98*, *101*)	(*24*, *137*, *200*, *201*)	(*24*, *138*, *202*, *203*)	(*20*)	(*20*, *114*, *204*)	(*23*, *105*, *106*)	1.0–21 (*205*)
Rhinovirus	(*9*, *27*)	(*26*, *28*)	–	(*26*–*28*)	–	(*27*)	(*9*)	1.2–2.7 (*205*)
Measles virus	(*16*)	(*16*)	–	–	(*17*)	(*17*)	(*16*)	12–18 (*206*)
Respiratory syncytial virus (RSV)	(*102*)	(*25*)	–	(*25*)	–	–	(*25*)	0.9–21.9 (*205*)

is dominated by short-range aerosol inhalation at most distances within 2 m of the infectious person, and droplets are only dominant when individuals are within 0.2 m when talking or 0.5 m when coughing (*50*). Anecdotal observations of measles virus (*16*–*18*) and *Mycobacterium tuberculosis* (*51*, *52*) infection in close proximity, previously attributed solely to droplets, include transmission by aerosols at short range. Further studies are warranted for respiratory diseases whose transmission has previously been characterized as droplet driven because it is plausible that airborne transmission is important or even dominant for most of them.

Early in the COVID-19 pandemic, it was assumed that droplets and fomites were the main transmission routes on the basis of the relatively low basic reproduction number (R_0) compared with that of measles (*53*–*55*) (Table 1). R_0 is the average number of secondary infections caused by a primary infected individual in a homogeneously susceptible population. This argument was built on a long-standing belief that all airborne diseases must be highly contagious. However, there is no scientific basis for such an assumption because airborne diseases exhibit a range of R_0 values that cannot be well represented by a single average value, which depends on numerous factors. For example, tuberculosis (R_0, 0.26 to 4.3) is an obligate airborne bacterial infection (*56*), but it is less transmissible than COVID-19 (R_0, 1.4 to 8.9) (*57*–*59*). The factors affecting airborne transmission include viral load in different-sized respiratory particles, the stability of the virus in aerosols, and the dose-response relationship for each virus (the probability of infection given exposure to a certain number of virions through a particular exposure route). Moreover, R_0 is an average, and COVID-19 is greatly overdispersed, meaning that, under certain conditions, it can be highly contagious. Epidemiological studies have found that 10 to 20% of infected individuals account for 80 to 90% of subsequent infections for SARS-CoV-2, highlighting the heterogeneity in secondary attack rates (the proportion of exposed individuals who become infected) (*60*–*63*).

A growing body of research on COVID-19 provides abundant evidence for the predominance of airborne transmission of SARS-CoV-2. This route dominates under certain environmental conditions, particularly indoor environments that are poorly ventilated (*6*, *34*, *35*, *41*, *42*, *45*, *50*, *64*–*68*), an observation that implicates solely aerosols because only aerosols—and not large droplets or surfaces—are affected by ventilation. Moreover, the marked difference between rates of indoor and outdoor transmission can only be explained by airborne transmission, because large droplets, whose trajectories are affected by gravitational settling but not ventilation, behave identically in both settings (*69*). Various combinations of epidemiological analyses; airflow model simulations; tracer experiments; and analysis and modeling of superspreading events in restaurants (*36*), in meatpacking plants (*70*), on a cruise ship (*71*), during singing at a choir rehearsal (*64*), and the long-distance transmission at a church (*72*) all implicate aerosols as the most likely mode of transmission over fomites and droplets. It is highly unlikely that most people at any of these events all touch the same contaminated surface or are exposed to droplets produced from the cough or sneeze of an infectious person at close range and encounter sufficient virus load to cause infection. However, the one common factor for all people at these indoor events is the shared air they inhale in the same room. Commonalities among superspreading events include indoor settings, crowds, exposure durations of 1 hour or more, poor ventilation, vocalization, and lack of properly worn masks (*36*). Given that droplet transmission dominates only when individuals are within 0.2 m when talking (*50*) and that transmission of SARS-CoV-2 through contaminated surfaces is less likely (*73*–*75*), superspreading events can only be explained by including aerosols as a mode of transmission.

To establish effective guidance and policies for protecting against airborne transmission of respiratory viruses, it is important to better understand the mechanisms involved. For airborne transmission to occur, aerosols must be generated, transported through air, inhaled by a susceptible host, and deposited in the respiratory tract to initiate infection. The virus must retain its infectivity throughout these processes. In this Review, we discuss the processes involved in the generation, transport, and deposition of virus-laden aerosols, as well as the important parameters that influence these processes, which are critical to informing effective infection control measures (Fig. 1).

Generation of virus-laden aerosols

Expiratory activities produce aerosols from different sites in the respiratory tract through distinct mechanisms. Aerosols produced by activities such as breathing, speaking, and coughing exhibit different aerosol size distributions and airflow velocities (*76*, *77*), which in turn govern the types and loads of viruses that each aerosol particle may carry, the residence time in air, the distance traveled, and ultimately the deposition sites in the respiratory tract of a person who inhales them (*78*). Aerosols released by an infected individual may contain viruses (*39*, *79*–*81*) as well as electrolytes, proteins, surfactants, and other components in the fluid that lines respiratory surfaces (*82*, *83*) (Fig. 2).

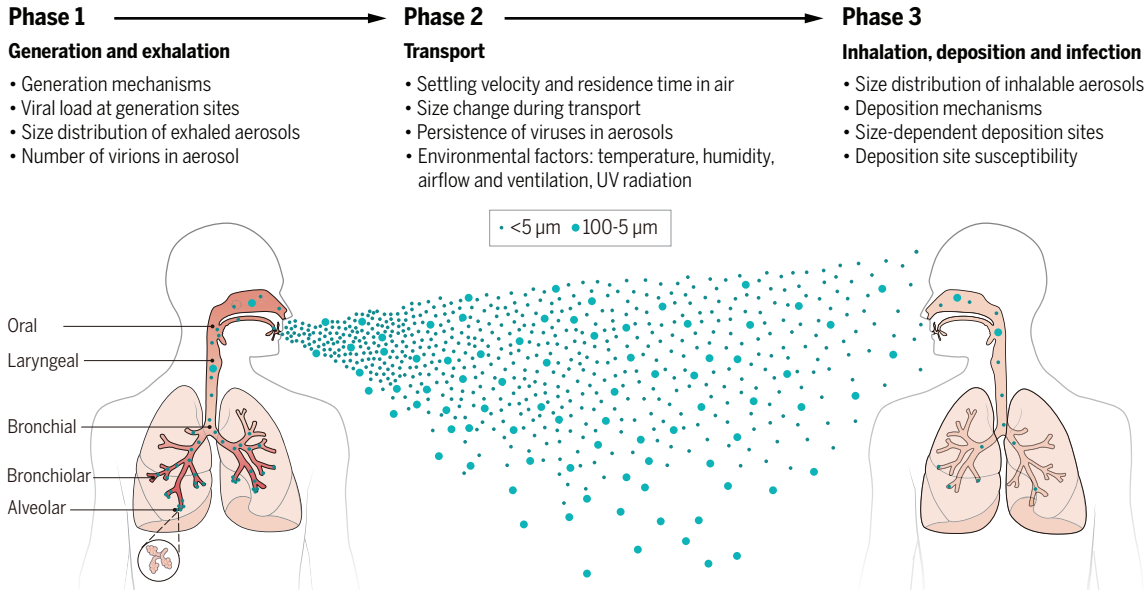


Fig. 1. Airborne transmission of respiratory viruses. Phases involved in the airborne transmission of virus-laden aerosols include (i) generation and exhalation; (ii) transport; and (iii) inhalation, deposition, and infection. Each phase is influenced by a combination of aerodynamic, anatomical, and environmental factors. (The sizes of virus-containing aerosols are not to scale.)

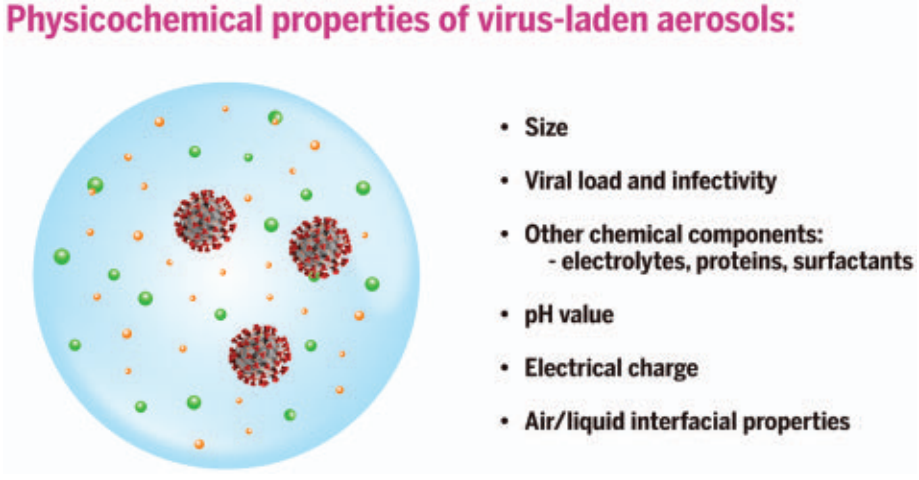


Fig. 2. Physicochemical properties of virus-laden aerosols. The behavior and fate of virus-laden aerosols are inherently governed by their characteristic properties, including physical size, viral load, infectivity, other chemical components in the aerosol, electrostatic charge, pH, and the air-liquid interfacial properties.

Sites of aerosol formation

Respiratory aerosols can be classified into alveolar, bronchiolar, bronchial, laryngeal, and oral aerosols, according to the sites where they are produced (*3*, *84*, *85*). Bronchiolar aerosols are formed during normal breathing (*3*). During exhalation, the liquid film lining the lumenal surfaces of the bronchioles ruptures to produce small aerosols. Such aerosols are generated by shear forces that destabilize the air-liquid or air-mucous interface. Respiratory airflows are often turbulent under high airflow velocities, particularly in the large lumens of the upper airways, which transition to laminar flow in the bronchi and bronchioles (*76*, *86*–*88*). Laryngeal aerosols are generated through vocal fold vibrations during vocalization (*3*). The apposition of vocal folds forms liquid bridges, which burst into aerosols during exhalation. By contrast, droplets (>100 μm) are primarily produced from saliva in the oral cavity (*3*). Aerosol emission rates increase with airflow velocity and speech volume during activities such as singing and shouting (*9*, *89*, *90*).

Number and size distributions

The size of exhaled aerosols is one of the most influential properties governing their fate, because size not only determines their aerodynamic characteristics but also their deposition dynamics and the site of infection. Size distributions of respiratory aerosols have been investigated since the 1890s using various approaches, including optical microscopy, high-speed photography, and, more recently, laser-based detection techniques (*1*, *2*, *91*). Early studies used measuring techniques and analytical methods that were unable to detect aerosols <5 μm (*1*, *92*), but current instruments, such as aerodynamic and scanning mobility particle sizing systems, have enabled the detection of smaller aerosols. Respiratory aerosols produce a multimodal size distribution, with peaks around 0.1 μm, 0.2 to 0.8 μm, 1.5 to 1.8 μm, and 3.5 to 5.0 μm, each representing a different generation site, production process, and expiratory activity (*2*, *8*, *9*, *85*, *91*, *93*). The smaller the modal size, the deeper the aerosols originate in the respiratory tract. A larger mode centered at 145 μm for talking and 123 μm for coughing originates mainly from the oral cavity and lips (*3*). In terms of number, the majority of exhaled aerosols are <5 μm, and a large fraction are <1 μm for most respiratory activities, including those produced during breathing, talking, and coughing (*8*, *9*). Overall, speech produces 100 to 1000 times the number of aerosols <100 μm in size for every droplet that is >100 μm (*3*).

Normal breathing has been shown to release up to 7200 aerosol particles per liter of exhaled air (*9*, *93*). The number of virus-laden aerosols expelled by individuals while breathing varies widely between individuals and depends on disease stage, age, body mass index,

and preexisting health conditions (*94*, *95*). Children generally produce fewer virus-laden aerosols than adults because their lungs are still developing and have fewer bronchioles and alveoli in which aerosols can form (*96*). The processes involved in aerosol formation, particularly the properties of fluid lining the airways that affect its propensity to break up to form aerosols, plays a crucial role in the number of aerosols exhaled (*94*). One study showed that 1 min of speaking may produce at least 1000 aerosols (*97*). Although coughing can produce more aerosols in a short period of time, it is much more sporadic than continuous breathing and speaking, especially for infected individuals who display no clinical symptoms. Therefore, breathing, speaking, and other continuous vocalization by infected individuals will likely release more total virus-laden aerosols overall than less-frequent coughing.

Viral content of aerosols

The viral load of aerosols is a key factor in determining the relative contribution of airborne transmission. However, sampling and detecting airborne viruses is challenging because of their low concentrations in air and susceptibility to destruction and inactivation during sampling. Air samples are often analyzed for the presence of viral genomes by quantitative polymerase chain reaction (qPCR) or quantitative reverse transcription PCR (qRT-PCR) methods, which are highly sensitive. Nevertheless, the presence of genetic material alone does not indicate whether the virus is infectious. The viability of viruses depends on the integrity and function of their genomic material, nucleoprotein, capsid, and/or envelope. Although some studies have tried and failed to culture viruses from air, the use of more gentle methods, such as a liquid condensation collection device, has enabled the detection of numerous viable respiratory viruses, including influenza viruses and SARS-CoV-2 in aerosols (*35*, *40*, *98*).

Many viruses have been isolated from breath and indoor air samples, including adenovirus (*29*, *99*), coxsackievirus (*100*), influenza viruses (*22*, *23*, *98*, *101*), rhinovirus (*9*, *26*–*28*), measles virus (*16*, *17*), RSV (*25*, *102*), SARS-CoV (*31*), MERS-CoV (*32*, *103*), and SARS-CoV-2 (*34*, *35*, *40*–*44*) (Table 1). The concentration of SARS-CoV-2 in the air of a hospital room with two COVID-19 patients was between 6 and 74 $TCID_{50}$ per liter (median tissue culture infectious dose per liter) (*35*). The distribution of virions across different sizes of aerosol particles is related to their site of generation, the production mechanism, and the severity of infection at the generation site, which varies among different viruses (*104*). It is commonly assumed that viral concentrations in clinical samples (e.g., sputum or saliva) translate directly to the concentration in droplets and aerosols generated from respiratory fluid—i.e., that viral load scales with the initial volume of droplets and aerosols (*50*, *55*, *71*). However, size-segregated samples of aerosols collected in the exhaled breath of individuals infected with influenza A or B viruses, parainfluenza virus, coronaviruses, hRV, or RSV and air collected in various settings show that viruses are enriched in smaller aerosols (*10*). In samples collected from influenza patients while breathing, talking, and/or coughing, more than half of the viral RNA was found in aerosols <4 to 5 μm (*23*, *104*, *105*). A study of several respiratory viruses found viral RNA more commonly in small (<5 μm) than in large aerosols (*39*). The distribution of influenza virus and RSV in ambient aerosols measured in a medical clinic revealed that 42% of influenza A virus RNA, but only 9% of RSV RNA, was in aerosols ≤4 μm (*102*). In a study that collected aerosols in a health clinic, childcare center, and airplanes, more than half of influenza A virus RNA was found in aerosols <2.5 μm (*106*). A study found that a subset of COVID-19 patients release up to 10^5 to 10^7 SARS-CoV-2 genome copies per hour in exhaled breath, whereas others do not exhale detectable virus (*107*). Large interpersonal variability in both the number of aerosols produced and their viral load may contribute to overdispersion in COVID-19 transmission, a crucial component in superspreading events (*108*).

Although infectious viruses are enriched in small aerosols, the dose-response relationship that governs the probability of infection given exposure to a certain number of virions, remains to be determined. In a susceptible host, the minimum infectious dose varies on the basis of virus type and deposition site within the respiratory tract, such that the inhalation of smaller aerosols that deposit deeper in the lungs could require less virus to initiate infection. Studies on influenza virus have shown that the dose required to initiate infection in humans, in terms of plaque-forming units (PFU), is, for the inhalation of aerosols, about a hundredth the size of the dose for intranasal inoculation (*101*). Improved characterization of the viral load and distribution of infectious virions in individual aerosols as a function of particle size, for different people and stages of disease, will greatly contribute to our understanding of airborne transmission of respiratory viruses.

Virus-laden aerosols in the environment

The physical characteristics of aerosols affect their transport in air. The initial velocity of respiratory aerosols depends on how they are generated within and released from the respiratory tract; for example, coughing produces droplets and aerosols released at higher velocities than speaking (*109*). Aerosol transport is controlled by a combination of airflow and environmental properties and by the physical characteristics of the aerosols themselves. Aerosols may diverge from streamlines as a result of inertia, Brownian motion, and external

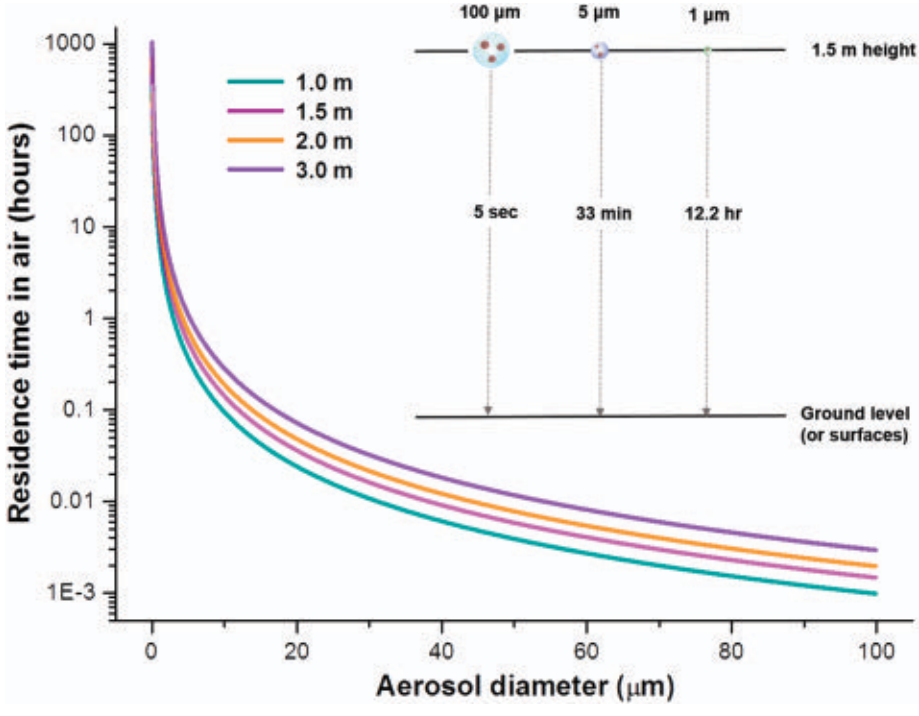


Fig. 3. How long can aerosols linger in air? Residence time of aerosols of varying size in still air can be estimated from Stokes' law for spherical particles (*116*). For example, the time required for an aerosol of 100, 5, or 1 μm to fall to the ground (or surfaces) from a height of 1.5 m is 5 s, 33 min, or 12.2 hours, respectively.

forces including gravitational, electrophoretic, and thermophoretic forces. Such motions can also lead to removal from air by deposition on surfaces. The lifetime of viruses in air is a function of physical transport and biological inactivation, which are affected by environmental factors, such as temperature, humidity, and ultraviolet (UV) radiation.

The sizes of exhaled aerosols that remain airborne evolve over time as a result of evaporation, coagulation, and/or deposition. Evaporation of water from aqueous aerosols is normally described by the Hertz-Knudsen equation (*110*). However, because respiratory aerosols contain nonvolatile components including proteins, electrolytes, and other biological species, the evaporation rate is slower than that of pure water (*111*). During evaporation, aerosols are subject to changes in phase, morphology, viscosity, and pH, all of which have been studied in simulated but not actual respiratory aerosols (*83*, *112*). Changes in physical characteristics of aerosols will affect the transport and fate of any viruses they contain, and associated changes in chemical characteristics of aerosols can affect virus viability (*113*). The overall size distributions of virus-laden aerosols in air also evolve over time because larger aerosols are preferentially removed by sedimentation to the ground or other surfaces, causing the median of the distribution to shift toward smaller sizes (*114*).

The residence time of virus-laden aerosols in air is crucial in determining their range of spread. In the absence of other forces, the residence time of an aerosol of a specific size is related to its terminal settling velocity, u_p, resulting from a balance between the viscous drag force and the gravitational force, as described by Stokes' law for small particles subject to laminar flow (*115*, *116*)

$$u_p = \frac{d_p^2 g \rho_p C_c}{18\eta}$$

where d_p is the diameter of the aerosol particle, g is gravitational acceleration, ρ_p is the density of the aerosol particle, C_c is the Cunningham slip correction factor accounting for the reduced air resistance caused by slippage when the particle size becomes comparable to the mean free path of gas molecules, and η is the dynamic viscosity of air.

The settling time for aerosols of a specific size to reach the ground can thus be estimated on the basis of an assumption that the surrounding air is at rest (Fig. 3). In still air, a 5-μm aerosol takes 33 min to settle to the ground from a height of 1.5 m, whereas a 1-μm aerosol can remain suspended in air for >12 hours (*116*). However, in most realistic environments, the velocity of the surrounding airflow should be taken into consideration. Additionally, when respiratory aerosols are exhaled, these particles are contained in an exhaled humid plume with its own speed and trajectory, which also play a role in determining the final reachable distance and direction (*86*). The distance that virus-laden aerosols travel depends on aerosol size, initial velocity of the flow carrying them, and other environmental conditions, such as outdoor wind speed or indoor air currents induced by natural ventilation or heating, ventilation, and air conditioning (HVAC) systems (*117*, *118*). The concentration of exhaled aerosols is highest close to the source (i.e., the infectious individual) and decreases with distance as the respiratory plume mixes with ambient air (*50*, *119*).

The trajectory and evaporation of exhaled aerosols generated during coughing and speaking have been studied with computational modeling (*117*, *120*). Large droplets tend to reach their maximum horizontal distances quickly and fall to the ground or surfaces within a few meters, whereas aerosols can remain suspended for many seconds to hours, travel long distances, and accumulate in air in poorly ventilated spaces (*117*). The multiphase nature of virus-laden aerosol flows greatly affects flow dynamics and how far aerosols travel, especially for exhalations with higher airflow velocities, such as in a cough (*121*).

Environmental factors that affect aerosol transmission

Survival of viruses in aerosols, also known as persistence, stability, or retention of infectivity, is commonly determined experimentally using a rotating drum, which allows the aerosols to remain suspended longer than in a stationary chamber. The decay of the virus can be described by first-order kinetics

$$C = C_o \times e^{-kt}$$

where C is the concentration of infectious viruses at time t, C_o is the initial concentration of infectious viruses, and k is the inactivation rate constant (*122*). The inactivation rate constant differs by virus and depends on a number of factors, including temperature, humidity, UV radiation, and chemical composition of the fluid from which the virus was aerosolized (*45*, *46*, *123*). This dependence, especially on respiratory fluid composition, makes it challenging to compare results across different studies. The time needed to reach 99.99% inactivation varies from hours to months (*124*). The decay rate can be quantified in terms of the half-life, which is ~1 to 3 hours for SARS-CoV and SARS-CoV-2 in laboratory-generated aerosols (*125*–*127*).

Temperature

Temperature is critical in mediating the survival and transmission of viruses in aerosols (*125*, *128*, *129*), likely by affecting the stability of the proteins, lipids, and genetic material that make up the virus. The upper respiratory tract is maintained at a few degrees cooler than the lungs (*130*), suggesting an enhanced replication capacity in the upper respiratory tract (*131*). SARS-CoV (*132*), SARS-CoV-2 (*133*), and influenza virus (*134*) are more stable at lower temperatures, possibly because of slower decay rates (as governed by the Arrhenius equation) and stronger ordering of phospholipids for enveloped viruses. Epidemiological evidence and animal studies suggest that the transmission of respiratory viruses known to infect the upper airways is favored at lower temperatures (*128*, *135*).

Relative humidity

By modulating the evaporation rate and equilibrium size of aerosols, relative humidity (RH) affects their transport and the viability of viruses they contain (*113*, *114*, *129*). Respiratory aerosols undergo evaporation upon release from the respiratory tract into ambient air as they transition from a saturated environment to lower RH. The evaporation process is expected to take seconds (*114*, *136*). At lower ambient RH, evaporation occurs more quickly and equilibrates at a smaller equilibrium size (*136*). At RH below ~80%, respiratory aerosols reach a final diameter that is 20 to 40% of the original size (*129*).

The seasonality of cases of influenza virus, human coronaviruses that cause common colds, RSV, and others has been at least partially attributed to RH (*134*). The sensitivity of a virus to RH may be influenced by RH-related effects on virus persistence in the environment and/or immune defenses. Mucociliary clearance is not as efficient at low RH (*134*). Animal studies have shown that influenza virus transmission is favored at low RH (*135*, *137*); however, a study of the 2009 pandemic influenza A virus (H1N1) in more physiologically realistic medium reported that the virus remained highly stable and infectious over a broad RH range between 20 and 100% (*138*). A study investigated the sensitivity of 11 airborne viruses to RH and found that although some RNA viruses survived best at low RH, other viruses survived better at high RH (*139*). The relationship between RH and virus viability in droplets and aerosols is characteristic to the virus, modulated by both the intrinsic physicochemical properties of the virus and its surrounding environment (*113*, *129*, *139*) (Fig. 2).

UV radiation

Irradiation with UV light has long been established as an effective approach to inactivate airborne viruses, including influenza virus (*127*, *140*), SARS-CoV, and other human coronaviruses (*141*). UV radiation rapidly inactivates SARS-CoV-2 in bulk culture medium (*142*) and in aerosols (*47*) at wavelengths found in ground-level sunlight. UV radiation damages genetic

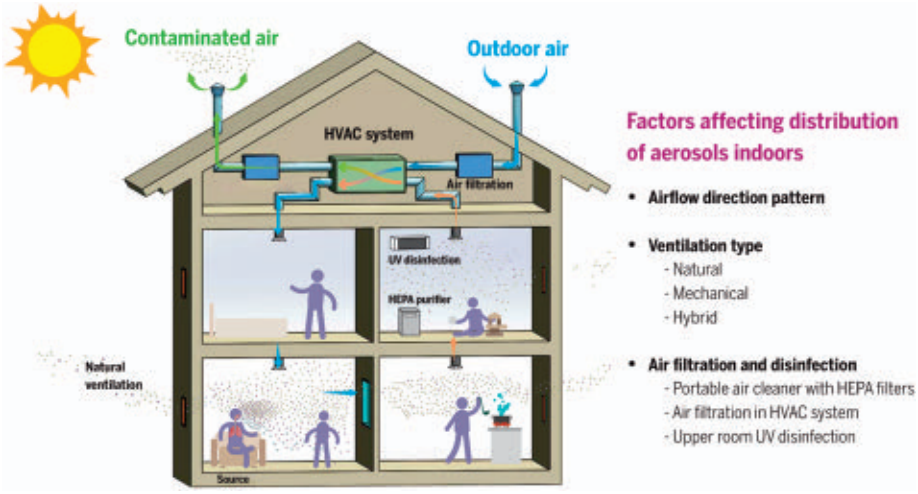


Fig. 4. Factors affecting indoor airborne transmission. Whereas the motion of large droplets is predominantly governed by gravity, the movement of aerosols is more strongly influenced by airflow direction and pattern, type of ventilation, and air filtration and disinfection.

material, leading to inactivation of the virus (*143*). Nevertheless, caution must be taken during operation of UV disinfection lamps to avoid direct eye and skin contact.

Airflow, ventilation, and filtration

Airflow strongly influences the transport of virus-laden aerosols (*81*) in contrast to droplets, which are rapidly deposited because of gravity. Aerosols in exhaled air tend to rise because the exhaled air is warmer than the environment (*50*), and their trajectories can also be influenced by the body's thermal plume (*81*). Greater airflow outdoors contributes to greater dispersion, whereas indoors the airflow is restricted by the surrounding walls and ceiling. Ventilation rate and airflow patterns play an important role in airborne transmission of viruses in indoor environments (*144*–*146*). A study of rhinovirus transmission showed that a low ventilation rate increases the risk of exposure to virus-laden aerosols indoors (*27*, *28*). An outbreak of COVID-19 in a high-rise apartment building occurred along vertically aligned units that were connected by a single air duct, demonstrating the risk of airborne transmission associated with shared air (*147*). Improving ventilation rates to reduce the carbon dioxide levels in under-ventilated buildings from 3200 parts per million (ppm) to 600 ppm (corresponding to an estimated increase of ventilation rate from 1.7 liters per second per person to 24 liters per second per person) has been shown to reduce the secondary attack rate of tuberculosis to zero (*146*).

The airflow in indoor environments is mediated by the design and operational status of ventilation systems, including the type of ventilation system (whether natural with open windows and doors, mechanical with blowers, or a hybrid of these), airflow patterns, air change rate, and supplementary systems such as air filtration (*145*, *148*) (Fig. 4). The WHO has recently recommended a ventilation rate of 10 liters per second per person (*149*). Proper placement of portable high-efficiency particulate air (HEPA) purifiers, which are capable of removing ≥99.97% of aerosol particles ≥0.3 μm, is also effective in reducing exposure of infectious aerosols, especially when combined with ventilation and universal masking (*150*–*152*). Although ventilation and filtration help to remove virus-laden aerosols, they must be implemented correctly to reduce the spread and risk of aerosol inhalation (*93*, *151*). A study quantitatively assessed the risk of airborne transmission of COVID-19 by asymptomatic individuals in elevator, classroom, and supermarket settings by combining in situ measurements and computational fluid dynamics (CFD) simulations, showing that inappropriate ventilation may create hotspots with risks much higher than in other room locations (*93*). Additionally, the physical plexiglass barriers designed to block droplet spray from coughs and sneezes in indoor spaces can impede the airflow and even trap higher concentrations of aerosols in the breathing zone and has been shown to increase transmission of SARS-CoV-2 (*153*).

The risk of airborne infection and correlation with ventilation rate can be assessed by a box model of virus transport and the Wells-Riley infection model (*17*, *64*)

$$P = \frac{N}{S} = 1 - e^{-Iqpt/Q}$$

where P is the probability of infection, N is the number of confirmed infection cases, S is the number of susceptible cases, I is number of infectors, q is the quanta (infectious dose) generation rate (quanta per hour), p is the pulmonary ventilation rate of susceptible individual (cubic meters per second), t is the exposure time (hours), and Q is the room ventilation rate (cubic meters per second). A model using the Wells-Riley method was applied to a large community outbreak of COVID-19 in a choir practice with one index case known to be symptomatic that led to 53 cases among 61 members in attendance (87% secondary attack rate), which concluded that poor ventilation along with a crowded venue, loud vocalization, and long duration all contributed to the high secondary attack rate (*64*). The choir practice had limited face-to-face interaction and strong attention on hand disinfection, which allowed major contributions from fomite or droplet transmission to be ruled out (*64*). Research is needed to establish minimum acceptable ventilation rates under different conditions and the effect of ventilation type on the risk of transmission.

Deposition of virus-laden aerosols

Once inhaled, virus-laden aerosols may deposit in the respiratory tract of a potential host. The size of aerosols is again central to determining the deposition site, although numerous anatomical, physiological, and aerodynamic factors (including the airway anatomical structure, breathing patterns, aerosol transport aerodynamics in the respiratory tract, and the physicochemical properties of inhaled aerosols) also affect the deposition pattern. Infection may be initiated at the deposition site if the virus remains infectious and appropriate receptors are present.

Aerosols up to 100 μm can be inhaled. Depending on their size, they deposit in different regions of the respiratory tract, based on one of several key mechanisms, including inertial impaction, gravitational sedimentation, Brownian diffusion, electrostatic precipitation, and interception (*154*, *155*) (Fig. 5A). Upon inhalation, the size of inhaled aerosols may increase as a result of hygroscopic growth in the nearly saturated respiratory tract (*156*). The International Committee for Radiological Protection (ICRP) has developed a model, based on human lung architecture, that quantifies deposition efficiency as a function of aerosol size (*157*) (Fig. 5B). Aerosols >5 μm deposit primarily in the nasopharyngeal region (87 to 95%), mainly through inertial impaction and gravitational sedimentation (*115*); although aerosols <5 μm also deposit there, they also may penetrate more deeply into the lungs and deposit in the alveolar lumen (*115*, *157*, *158*). Brownian diffusion is the dominant deposition mechanism of inhaled particles <0.1 μm in the bronchiolar and alveolar regions (*78*, *116*, *159*). Aerosols that carry natural electrostatic charge may be attracted to the airway walls (*160*). Provided a cellular receptor is present at the deposition site, infection may be initiated. The infection efficiency is further governed by

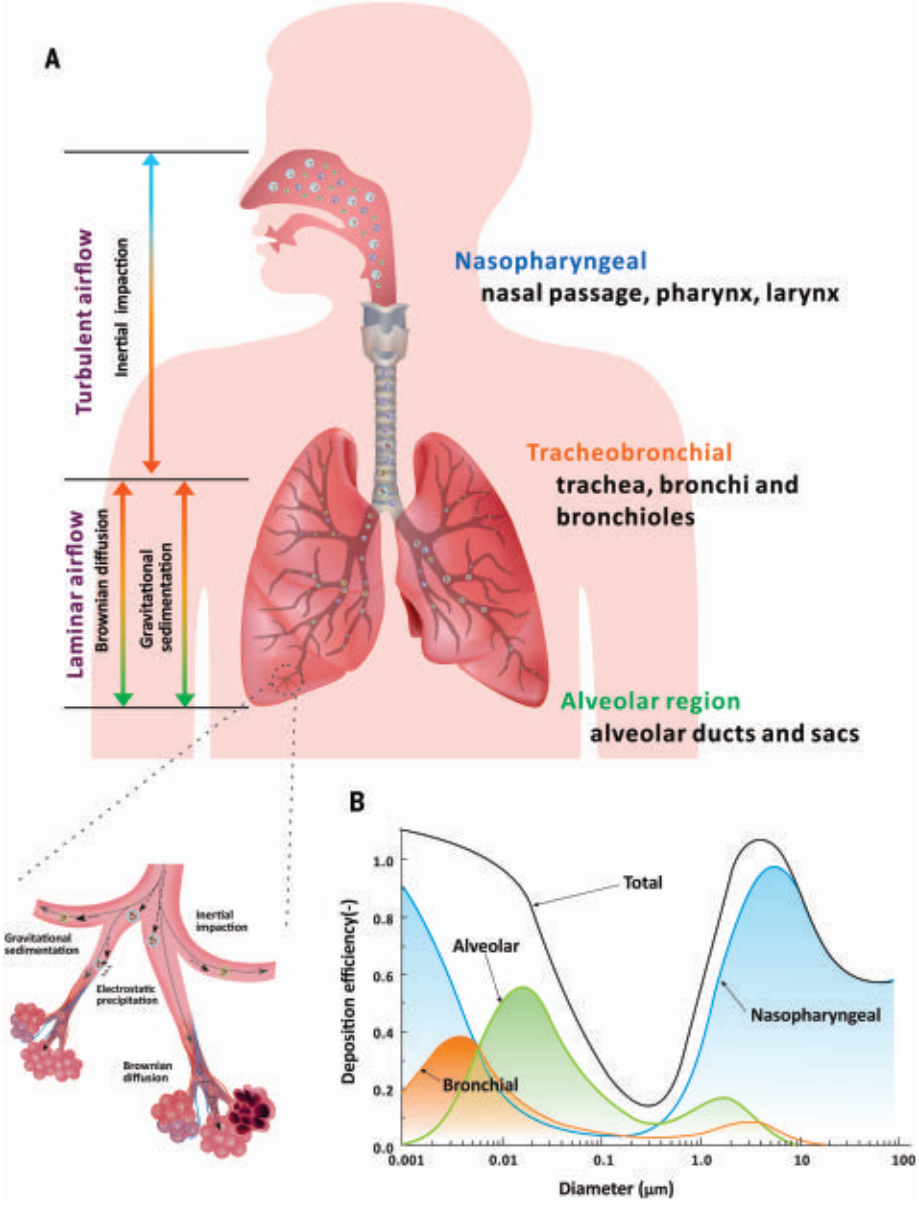


Fig. 5. Size-dependent aerosol deposition mechanisms to sites in the respiratory tract. (**A**) Main deposition mechanisms and corresponding airflow regimes in different regions of the human respiratory tract. Large aerosols tend to deposit in the nasopharyngeal region as a result of inertial impaction, whereas small aerosols tend to deposit in the tracheobronchial and alveolar regions on the basis of gravitational sedimentation and Brownian diffusion. An enlarged view of tracheobronchial and alveolar regions illustrates the deposition mechanism. (**B**) The deposition efficiency of aerosols at different regions of the respiratory tract as a function of aerosol diameter based on the ICRP lung deposition model is shown (*116*). The majority of large aerosols deposit in the nasopharyngeal region; only aerosols that are sufficiently small can reach and deposit in the alveolar region.

the distribution of cellular receptors along the respiratory tract and the virus-host interaction.

Deposition of aerosols in diseased lungs may differ from that in normal lungs because of airway surface structure changes and obstruction by mucous (*161*). Changes in the surface properties of the respiratory epithelium in asthmatic airways and airway narrowing as a result of chronic obstructive pulmonary disease (COPD) alter the airflow and aerodynamic behaviors of inhaled aerosols, thus modifying their deposition dynamics and sites (*162*, *163*). Deposition is generally higher in patients with COPD than in healthy individuals; bronchial deposition is higher in patients with asthma and chronic bronchitis (*154*).

Because viruses are enriched in small aerosols (<5 μm), they can travel deeper into and be deposited in the lower respiratory tract. The viral load of SARS-CoV-2 has been reported to be higher and the virus persists longer in the lower respiratory tract compared with the upper respiratory tract (*164*, *165*). Initiation of an infection in the lower respiratory tract adds technical challenges in diagnosing patients because current screening commonly collects samples from the nasopharyngeal or oral cavity using swabs.

Discussion

Airborne transmission has long been an underappreciated route for contributing to the transmission of respiratory viral diseases, largely because of an insufficient understanding of the generation and transport processes of virus-laden aerosols as well as misattribution of anecdotal observations. The epidemiological evidence for the dominance of airborne spread of SARS-CoV-2 has increased over time and has become especially strong. First, the distinct difference between indoor and outdoor transmission cannot be explained by droplet transmission because gravity-driven droplets behave identically indoors and outdoors. The high frequency of indoor superspreading events relative to those outdoors points to the importance of airborne transmission (*63*). The demonstrated role of poor ventilation in transmission and superspreading clusters indoors is also only compatible with aerosols, because droplets and fomite transmission are not affected by ventilation. Long-range airborne transmission of SARS-CoV-2 has been observed in hotel quarantines in countries with very low transmission (*166*) and in a large church (*72*).

During the emergence of novel respiratory viruses, a more holistic approach that acknowledges all modes of transmission (airborne, droplet, and fomite) is needed to successfully mitigate risk and prevent spread. The requirement for direct evidence of infectiousness of sampled aerosols before acknowledging and adding controls to address airborne transmission leaves people at potential risk (*69*). When unburdened by conventional definitions of transmission routes, the available evidence for SARS-CoV-2, influenza virus, and other respiratory viruses is much more consistent with transmission by aerosols <100 μm rather than by rare, large droplets sprayed onto mucous membranes of people in very close proximity. Recent acknowledgement of airborne transmission of SARS-CoV-2 by the WHO (*48*) and US CDC (*49*) reinforces the necessity to implement protection against this transmission route at both short and long ranges.

Once the mechanisms leading to airborne transmission are fully understood—acknowledging that transmission by aerosols is largest at close range—it becomes clear there is an overlap in precautions and mitigation measures for both droplets and aerosols (such as distancing and masks), but extra considerations must be taken into account for mitigating aerosol transmission at both short and long ranges. These

include attention to ventilation, airflows, mask fit and type, air filtration, and UV disinfection, as well as distinguishing measures between indoor and outdoor environments. Although our knowledge is still increasing, enough is already known to add protective measures to better protect against airborne transmission of respiratory viruses, noting that "droplet precautions" are not replaced but instead expanded.

A high proportion of individuals infected with SARS-CoV-2 have no symptoms at the time of testing (*167*, *168*). About 20 to 45% of individuals infected with SARS-CoV-2 remained asymptomatic throughout the course of infection, whereas some infected individuals experienced a presymptomatic phase and began to develop symptoms several days after infection (*168*, *169*). The infectiousness of SARS-CoV-2 peaks two days before and extends to one day after symptom onset (*170*). High asymptomatic infection rates have also been reported for influenza virus and other respiratory virus infections (*171*–*173*). Although some studies suggest that airborne transmission is not an efficient route, particularly for asymptomatic and mildly symptomatic individuals who likely have low viral loads in their saliva (*55*), the viral load in presymptomatic individuals is comparable to that of symptomatic patients (*174*, *175*). It is important to implement controls that protect against exposure of infectious virus-laden aerosols produced when infected individuals without any symptoms speak, sing, or simply breathe. Because these individuals do not know they are infected, they generally continue to be involved in social activities, leading to airborne transmission.

Universal masking is an effective and economical way to block virus-laden aerosols (*67*). Model simulations show that masks effectively prevent asymptomatic transmission and reduce the total number of infected individuals as well as mortalities as a result of COVID-19 (*176*). It is crucial to optimize the allocation of masks (*177*). Surgical masks have been shown to reduce the release of influenza virus, seasonal human coronaviruses, and rhinovirus in aerosols <5 μm into the air by infected individuals by up to 100% (*104*, *178*), although for some individuals there was no reduction; and masks are more effective for limiting droplets (*179*). Masks made of combinations of different fabrics and/or multiple layers, when worn properly with no leaks, can block up to 90% of particles between 0.5 and 10 μm (*179*). Small gaps between the mask material and skin can lead to substantial decreases in the overall filtration efficiency. For aerosols <2.5 μm, filtration efficiency decreases by 50% for a relative leak area of 1% (*180*). A study compared the viral filtration efficiency of N95, surgical, and fabric masks using a model virus and found that the efficiency of N95 and some surgical masks exceeded 99%; all fabric masks tested were at least 50% efficient (*181*). The effectiveness of N95, surgical, and cotton masks in blocking SARS-CoV-2–containing aerosols has been investigated using manikins placed face-to-face. N95 respirators demonstrated the highest efficiency in blocking infectious SARS-CoV-2 (*182*). Almost all masks offer at least some protection, but they are not 100% effective. Transmission of SARS-CoV-2 has occurred in health care settings despite medical masks (designed for droplets not aerosols) and eye protection (*183*–*185*), which illustrates the need for proper personal protective equipment (PPE) and layering multiple interventions against airborne transmission, especially in high-risk indoor settings.

Health care facilities are more likely to accommodate patients infected with respiratory viruses. Thus, health care personnel should be provided with proper PPE to reduce airborne exposure. People occupying indoor spaces have increased potential to be exposed to high concentrations of virus-laden aerosols, especially in poorly ventilated and/or crowded indoor settings where virus-laden aerosols can readily accumulate (*93*). Preventive measures should be implemented at all times when traveling in airplanes, trains, buses, ships, and cruise ships, which have relatively small and enclosed air spaces where the ventilation may not always be optimal. Many studies indicate that the risk of airborne transmission in outdoor environments is substantially lower than indoor environments (*186*); however, the risk of transmission outdoors exists in close proximity situations, especially if talking, singing, or shouting over time. The risk of outdoor transmission may rise with increased lifetime and transmissibility of viruses, such as certain variants of SARS-CoV-2 (*187*, *188*). Aerosolization of virus-containing wastewater and hospital fecal discharges also poses potential outdoor exposure risks, which should not be underestimated (*189*).

Implementing effective ventilation systems reduces airborne transmission of infectious virus-laden aerosols. Strategies such as ensuring sufficient ventilation rates and avoiding recirculation are advised (*190*, *191*). Carbon dioxide sensors can be used as indicators of the build-up of exhaled air and serve as a simple way to monitor and optimize ventilation (*192*, *193*). Aerosol sensors can also be used to assess HEPA and HVAC aerosol filtration efficiencies, which are key to lowering infections caused by virus-laden aerosols. Assuring a minimum ventilation rate of 4 to 6 air changes per hour (ACH) and maintaining carbon dioxide levels below 700 to 800 ppm have been advised, although the ventilation type and airflow direction and pattern should also be taken into account (*148*, *194*). Increasing the efficiency of air filtration in HVAC systems, stand-alone HEPA purifiers, or implementing upper room UV disinfection systems can further reduce the concentrations of virus-laden aerosols (*47*, *127*, *140*, *141*, *195*).

Physical distancing, a mitigation put in place to address droplet transmission, is also effective in reducing the chances of aerosol inhalation because aerosol concentrations are much higher in close proximity to an infected individual (*50*). The WHO and many national public health agencies recommend maintaining physical distances of either 1 or 2 m. However, this distance is not sufficient to protect against aerosols that travel beyond this range. If large droplets dominated transmission, distancing alone would have effectively suppressed the transmission of SARS-CoV-2. As has been repeatedly shown in superspreading events, airborne transmission occurs in poorly ventilated rooms when occupants inhale infectious room air (*18*, *36*, *62*, *64*, *71*). Additionally, although distancing helps by moving people away from the most concentrated parts of respiratory plumes, distancing alone does not stop transmission and is not sufficient without accounting for other measures, such as ventilation and filtration, the number of people emitting infectious aerosols, and the amount of time spent in enclosed spaces (*196*). The unknown number of asymptomatic (including presymptomatic) infected individuals present in specific environmental settings is an additional challenge in respiratory disease control. Engineering measures to reduce aerosol concentrations through ventilation, filtration, and upper room UV disinfection remain critical strategies for reducing airborne transmission risks.

Despite the emerging recognition of airborne transmission of respiratory viruses, numerous issues require further exploration. For example, direct measurements are needed of the concentration of virus in aerosols and droplets as a function of size and their potential to initiate a new infection. The lifetime of viruses in aerosols of varying size requires systematic investigation. More studies are needed to quantify the relationship between viral dose delivered by aerosols and droplets and severity of infection; this relationship likely varies considerably for different viruses. It is also important to investigate whether the severity of disease correlates with the size and number of aerosols and the location in which they are deposited in the respiratory tract. Although more studies are needed, unequivocal evidence indicates that airborne transmission is a major pathway for the spread of SARS-CoV-2 and many other respiratory viruses. Additional precautionary measures must be implemented for mitigating aerosol transmission at both short and long ranges, with a major focus on ventilation, airflows, air filtration, UV disinfection, and mask fit. These interventions are critical strategies for helping end the current pandemic and preventing future outbreaks. It is

important to note that these proposed measures to improve indoor air quality will lead to long overdue improvements that have health benefits extending well beyond the COVID-19 pandemic.

REFERENCES AND NOTES

1. J. P. Duguid, The size and the duration of air-carriage of respiratory droplets and droplet-nuclei. *Epidemiol. Infect.* **44**, 471–479 (1946). doi: 10.1017/S0022172400019288; pmid: 20475760
2. L. Morawska *et al*., Size distribution and sites of origin of droplets expelled from the human respiratory tract during expiratory activities. *J. Aerosol Sci.* **40**, 256–269 (2009). doi: 10.1016/j.jaerosci.2008.11.002
3. G. R. Johnson *et al*., Modality of human expired aerosol size distributions. *J. Aerosol Sci.* **42**, 839–851 (2011). doi: 10.1016/j.jaerosci.2011.07.009
4. G. Scheuch, Breathing is enough: For the spread of influenza virus and SARS-CoV-2 by breathing only. *J. Aerosol Med. Pulm. Drug Deliv.* **33**, 230–234 (2020). doi: 10.1089/jamp.2020.1616; pmid: 32552296
5. W. F. Wells, On air-borne infection: Study II. Droplets and droplet nuclei. *Am. J. Epidemiol.* **20**, 611–618 (1934). doi: 10.1093/oxfordjournals.aje.a118097
6. The National Academies of Sciences, Engineering, and Medicine (NASEM), "Airborne transmission of SARS-CoV-2: A virtual workshop, 26 to 27 August 2020" (NASEM, 2020); www.nationalacademies.org/event/08-26-2020/airborne-transmission-of-sars-cov-2-a-virtual-workshop.
7. K. A. Prather *et al*., Airborne transmission of SARS-CoV-2. *Science* **370**, 303–304 (2020). doi: 10.1126/science.abf0521; pmid: 33020250
8. G. Zayas *et al*., Cough aerosol in healthy participants: Fundamental knowledge to optimize droplet-spread infectious respiratory disease management. *BMC Pulm. Med.* **12**, 11 (2012). doi: 10.1186/1471-2466-12-11; pmid: 22436202
9. P. Fabian, J. Brain, E. A. Houseman, J. Gern, D. K. Milton, Origin of exhaled breath particles from healthy and human rhinovirus-infected subjects. *J. Aerosol Med. Pulm. Drug Deliv.* **24**, 137–147 (2011). doi: 10.1089/jamp.2010.0815; pmid: 21361786
10. K. P. Fennelly, Particle sizes of infectious aerosols: Implications for infection control. *Lancet Respir. Med.* **8**, 914–924 (2020). doi: 10.1016/S2213-2600(20)30323-4; pmid: 32717211
11. C. A. E. Winslow, *Conquest of Epidemic Disease* (Princeton Univ. Press, 1943).
12. B. Rush, *The Works of Thomas Sydenham, M.D., On Acute and Chronic Diseases: With Their Histories and Modes of Cure* (Benjamin & Thomas Kite, 1809).
13. C. V. Chapin, *The Sources and Modes of Infection* (Wiley, 1910).
14. K. Han *et al*., Lack of airborne transmission during outbreak of pandemic (H1N1) 2009 among tour group members, China, June 2009. *Emerg. Infect. Dis.* **15**, 1578–1581 (2009). doi: 10.3201/eid1510.091013; pmid: 19861048
15. A. Bak *et al*., SARS-CoV-2 routes of transmission and recommendations for preventing acquisition: Joint British Infection Association (BIA), Healthcare Infection Society (HIS), Infection Prevention Society (IPS) and Royal College of Pathologists (RCPath) guidance. *J. Hosp. Infect.* **114**, 79–103 (2021). doi: 10.1016/j.jhin.2021.04.027; pmid: 33940093
16. W. E. Bischoff *et al*., Detection of measles virus RNA in air and surface specimens in a hospital setting. *J. Infect. Dis.* **213**, 600–603 (2016). doi: 10.1093/infdis/jiv465; pmid: 26386428
17. E. C. Riley, G. Murphy, R. L. Riley, Airborne spread of measles in a suburban elementary school. *Am. J. Epidemiol.* **107**, 421–432 (1978). doi: 10.1093/oxfordjournals.aje.a112560; pmid: 665658
18. A. B. Bloch *et al*., Measles outbreak in a pediatric practice: Airborne transmission in an office setting. *Pediatrics* **75**, 676–683 (1985). pmid: 3982900
19. R. Tellier, Review of aerosol transmission of influenza A virus. *Emerg. Infect. Dis.* **12**, 1657–1662 (2006). doi: 10.3201/eid1211.060426; pmid: 17283614
20. B. J. Cowling *et al*., Aerosol transmission is an important mode of influenza A virus spread. *Nat. Commun.* **4**, 1935 (2013). doi: 10.1038/ncomms2922; pmid: 23736803
21. R. Tellier, Aerosol transmission of influenza A virus: A review of new studies. *J. R. Soc. Interface* **6**, S783–S790 (2009). doi: 10.1098/rsif.2009.0302.focus; pmid: 19773292
22. W. E. Bischoff, K. Swett, I. Leng, T. R. Peters, Exposure to influenza virus aerosols during routine patient care. *J. Infect. Dis.* **207**, 1037–1046 (2013). doi: 10.1093/infdis/jis773; pmid: 23372182
23. J. Yan *et al*., Infectious virus in exhaled breath of symptomatic seasonal influenza cases from a college community. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1081–1086 (2018). doi: 10.1073/pnas.1716561115; pmid: 29348203
24. F. Koster *et al*., Exhaled aerosol transmission of pandemic and seasonal H1N1 influenza viruses in the ferret. *PLOS ONE* **7**, e33118 (2012). doi: 10.1371/journal.pone.0033118; pmid: 22509254
25. H. Kulkarni *et al*., Evidence of respiratory syncytial virus spread by aerosol. Time to revisit infection control strategies? *Am. J. Respir. Crit. Care Med.* **194**, 308–316 (2016). doi: 10.1164/rccm.201509-18330C; pmid: 26890617
26. E. C. Dick, L. C. Jennings, K. A. Mink, C. D. Wartgow, S. L. Inhorn, Aerosol transmission of rhinovirus colds. *J. Infect. Dis.* **156**, 442–448 (1987). doi: 10.1093/infdis/156.3.442; pmid: 3039011
27. T. A. Myatt *et al*., Detection of airborne rhinovirus and its relation to outdoor air supply in office environments. *Am. J. Respir. Crit. Care Med.* **169**, 1187–1190 (2004). doi: 10.1164/rccm.200306-760OC; pmid: 14754759
28. T. A. Myatt, S. L. Johnston, S. Rudnick, D. K. Milton, Airborne rhinovirus detection and effect of ultraviolet irradiation on detection by a semi-nested RT-PCR assay. *BMC Public Health* **3**, 5 (2003). doi: 10.1186/1471-2458-3-5; pmid: 12525263
29. C.-C. Tseng, L.-Y. Chang, C.-S. Li, Detection of airborne viruses in a pediatrics department measured using real-time qPCR coupled to an air-sampling filter method. *J. Environ. Health* **73**, 22–28 (2010). pmid: 21133312
30. I. T. S. Yu *et al*., Evidence of airborne transmission of the severe acute respiratory syndrome virus. *N. Engl. J. Med.* **350**, 1731–1739 (2004). doi: 10.1056/NEJMoa032867; pmid: 15102999
31. T. F. Booth *et al*., Detection of airborne severe acute respiratory syndrome (SARS) coronavirus and environmental contamination in SARS outbreak units. *J. Infect. Dis.* **191**, 1472–1477 (2005). doi: 10.1086/429634; pmid: 15809906
32. S. H. Kim *et al*., Extensive viable Middle East respiratory syndrome (MERS) coronavirus contamination in air and surrounding environment in MERS isolation wards. *Clin. Infect. Dis.* **63**, 363–369 (2016). doi: 10.1093/cid/ciw239; pmid: 27090992
33. J. S. Kutter *et al*., SARS-CoV and SARS-CoV-2 are transmitted through the air between ferrets over more than one meter distance. *Nat. Commun.* **12**, 1653 (2021). doi: 10.1038/s41467-021-21918-6; pmid: 33712573
34. J. L. Santarpia *et al*., The Infectious Nature of Patient-Generated SARS-CoV-2 Aerosol. medRxiv 2020.07.13.20041632 [Preprint] (2020). doi: 10.1101/2020.07.13.20041632
35. J. A. Lednicky *et al*., Viable SARS-CoV-2 in the air of a hospital room with COVID-19 patients. *Int. J. Infect. Dis.* **100**, 476–482 (2020). doi: 10.1016/j.ijid.2020.09.025; pmid: 32949774
36. Y. Li *et al*., Probable airborne transmission of SARS-CoV-2 in a poorly ventilated restaurant. *Build. Environ.* **196**, 107788 (2021). doi: 10.1016/j.buildenv.2021.107788; pmid: 33746341
37. S. F. Sia *et al*., Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. *Nature* **583**, 834–838 (2020). doi: 10.1038/s41586-020-2342-5; pmid: 32408338
38. J. Shi *et al*., Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* **368**, 1016–1020 (2020). doi: 10.1126/science.abb7015; pmid: 32269068
39. J. Gralton, E. R. Tovey, M.-L. McLaws, W. D. Rawlinson, Respiratory virus RNA is detectable in airborne and droplet particles. *J. Med. Virol.* **85**, 2151–2159 (2013). doi: 10.1002/jmv.23698; pmid: 23959825
40. J. A. Lednicky *et al*., Isolation of SARS-CoV-2 from the air in a car driven by a COVID patient with mild illness. *Int. J. Infect. Dis.* **108**, 212–216 (2021). doi: 10.1016/j.ijid.2021.04.063; pmid: 33901650
41. Y. Liu *et al*., Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature* **582**, 557–560 (2020). doi: 10.1038/s41586-020-2271-3; pmid: 32340022
42. Z.-D. Guo *et al*., Aerosol and surface distribution of severe acute respiratory syndrome coronavirus 2 in hospital wards, Wuhan, China, 2020. *Emerg. Infect. Dis.* **26**, 1583–1591 (2020). doi: 10.3201/eid2607.200885; pmid: 32275497
43. P. Y. Chia *et al*., Detection of air and surface contamination by SARS-CoV-2 in hospital rooms of infected patients. *Nat. Commun.* **11**, 2800 (2020). doi: 10.1038/s41467-020-16670-2; pmid: 32472043
44. J. L. Santarpia *et al*., Aerosol and surface contamination of SARS-CoV-2 observed in quarantine and isolation care. *Sci. Rep.* **10**, 12732 (2020). doi: 10.1038/s41598-020-69286-3; pmid: 32728118
45. N. van Doremalen *et al*., Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* **382**, 1564–1567 (2020). doi: 10.1056/NEJMc2004973; pmid: 32182409
46. S. J. Smither, L. S. Eastaugh, J. S. Findlay, M. S. Lever, Experimental aerosol survival of SARS-CoV-2 in artificial saliva and tissue culture media at medium and high humidity. *Emerg. Microbes Infect.* **9**, 1415–1417 (2020). doi: 10.1080/22221751.2020.1777906; pmid: 32496967
47. M. Schuit *et al*., Airborne SARS-CoV-2 is rapidly inactivated by simulated sunlight. *J. Infect. Dis.* **222**, 564–571 (2020). doi: 10.1093/infdis/jiaa334; pmid: 32525979
48. World Health Organization (WHO), "Coronavirus disease (COVID-19): How is it transmitted?" (2021); www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted.
49. U.S. Centers for Disease Control and Prevention (CDC), "Scientific brief: SARS-CoV-2 transmission" (2021); www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/sars-cov-2-transmission.html.
50. W. Chen, N. Zhang, J. Wei, H.-L. Yen, Y. Li, Short-range airborne route dominates exposure of respiratory infection during close contact. *Build. Environ.* **176**, 106859 (2020). doi: 10.1016/j.buildenv.2020.106859
51. R. L. Riley *et al*., Aerial dissemination of pulmonary tuberculosis a two-year study of contagion in a tuberculosis ward1. *Am. J. Epidemiol.* **70**, 185–196 (1959). doi: 10.1093/oxfordjournals.aje.a120069
52. E. A. Nardell, Transmission and institutional infection control of tuberculosis. *Cold Spring Harb. Perspect. Med.* **6**, a018192 (2015). doi: 10.1101/cshperspect.a018192; pmid: 26292985
53. M. Klompas, M. A. Baker, C. Rhee, Airborne transmission of SARS-CoV-2: Theoretical considerations and available evidence. *JAMA* **324**, 441–442 (2020). doi: 10.1001/jama.2020.12458; pmid: 32749495
54. J. Conly *et al*., Use of medical face masks versus particulate respirators as a component of personal protective equipment for health care workers in the context of the COVID-19 pandemic. *Antimicrob. Resist. Infect. Control* **9**, 126 (2020). doi: 10.1186/s13756-020-00779-6; pmid: 32762735
55. S. H. Smith *et al*., Aerosol persistence in relation to possible transmission of SARS-CoV-2. *Phys. Fluids* **32**, 107108 (2020). doi: 10.1063/5.0027844; pmid: 33154612
56. Y. Ma, C. R. Horsburgh Jr., L. F. White, H. E. Jenkins, Quantifying TB transmission: A systematic review of reproduction number and serial interval estimates for tuberculosis. *Epidemiol. Infect.* **146**, 1478–1494 (2018). doi: 10.1017/S0950268818001760; pmid: 29970199
57. Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, taaa021 (2020). doi: 10.1093/jtm/taaa021; pmid: 32052846
58. S. Sanche *et al*., High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26**, 1470–1477 (2020). doi: 10.3201/eid2607.200282; pmid: 32255761
59. C. R. MacIntyre, M. R. Ananda-Rajah, Scientific evidence supports aerosol transmission of SARS-COV-2. *Antimicrob. Resist. Infect. Control* **9**, 202 (2020). doi: 10.1186/s13756-020-00868-6; pmid: 33339522
60. R. Laxminarayan *et al*., Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* **370**, 691–697 (2020). doi: 10.1126/science.abd7672; pmid: 33154136
61. K. Sun *et al*., Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* **371**, eabe2424 (2021). doi: 10.1126/science.abe2424; pmid: 33234698
62. D. C. Adam *et al*., Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26**,

1714–1719 (2020). doi: 10.1038/s41591-020-1092-0; pmid: 32943787
63. D. Lewis, Superspreading drives the COVID pandemic - and could help to tame it. *Nature* **590**, 544–546 (2021). doi: 10.1038/d41586-021-00460-x; pmid: 33623168
64. S. L. Miller *et al*., Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event. *Indoor Air* **31**, 314–323 (2021). doi: 10.1111/ina.12751; pmid: 32979298
65. L. Morawska, D. K. Milton, It is time to address airborne transmission of coronavirus disease 2019 (COVID-19). *Clin. Infect. Dis.* **71**, 2311–2313 (2020). doi: 10.1093/cid/ciaa939; pmid: 32628269
66. E. L. Anderson, P. Turnham, J. R. Griffin, C. C. Clarke, Consideration of the aerosol transmission for COVID-19 and public health. *Risk Anal.* **40**, 902–907 (2020). doi: 10.1111/risa.13500; pmid: 32356927
67. K. A. Prather, C. C. Wang, R. T. Schooley, Reducing transmission of SARS-CoV-2. *Science* **368**, 1422–1424 (2020). doi: 10.1126/science.abc6197; pmid: 32461212
68. L. Morawska, J. Cao, Airborne transmission of SARS-CoV-2: The world should face the reality. *Environ. Int.* **139**, 105730 (2020). doi: 10.1016/j.envint.2020.105730; pmid: 32294574
69. T. Greenhalgh *et al*., Ten scientific reasons in support of airborne transmission of SARS-CoV-2. *Lancet* **397**, 1603–1605 (2021). doi: 10.1016/S0140-6736(21)00869-2; pmid: 33865497
70. J. Middleton, R. Reintjes, H. Lopes, Meat plants-a new front line in the covid-19 pandemic. *BMJ* **370**, m2716 (2020). doi: 10.1136/bmj.m2716; pmid: 32646892
71. P. Azimi, Z. Keshavarz, J. G. Cedeno Laurent, B. Stephens, J. G. Allen, Mechanistic transmission modeling of COVID-19 on the *Diamond Princess* cruise ship demonstrates the importance of aerosol transmission. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015482118 (2021). doi: 10.1073/pnas.2015482118; pmid: 33536312
72. A. L. Katelaris *et al*., Epidemiologic evidence for airborne transmission of SARS-CoV-2 during church singing, Australia, 2020. *Emerg. Infect. Dis.* **27**, 1677–1680 (2021). doi: 10.3201/eid2706.210465; pmid: 33818372
73. E. Goldman, Exaggerated risk of transmission of COVID-19 by fomites. *Lancet Infect. Dis.* **20**, 892–893 (2020). doi: 10.1016/S1473-3099(20)30561-2; pmid: 32628907
74. M. U. Mondelli, M. Colaneri, E. M. Seminari, F. Baldanti, R. Bruno, Low risk of SARS-CoV-2 transmission by fomites in real-life conditions. *Lancet Infect. Dis.* **21**, e112 (2021). doi: 10.1016/S1473-3099(20)30678-2; pmid: 33007224
75. A. K. Pitol, T. R. Julian, Community transmission of SARS-CoV-2 by surfaces: Risks and risk reduction strategies. *Environ. Sci. Technol. Lett.* **8**, 263–269 (2021). doi: 10.1021/acs.estlett.0c00966
76. M. Abkarian, S. Mendez, N. Xue, F. Yang, H. A. Stone, Speech can produce jet-like transport relevant to asymptomatic spreading of virus. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25237–25245 (2020). doi: 10.1073/pnas.2012156117; pmid: 32978297
77. L. Bourouiba, The fluid dynamics of disease transmission. *Annu. Rev. Fluid Mech.* **53**, 473–508 (2021). doi: 10.1146/annurev-fluid-060220-113712
78. R. J. Thomas, Particle size and pathogenicity in the respiratory tract. *Virulence* **4**, 847–858 (2013). doi: 10.4161/viru.27172; pmid: 24225380
79. S. A. Sattar, M. K. Ijaz, C. P. Gerba, Spread of viral infections by aerosols. *Crit. Rev. Environ. Control* **17**, 89–131 (1987). doi: 10.1080/10643388709388331
80. R. M. Jones, L. M. Brosseau, Aerosol transmission of infectious disease. *J. Occup. Environ. Med.* **57**, 501–508 (2015). doi: 10.1097/JOM.0000000000000448; pmid: 25816216
81. J. Wei, Y. Li, Airborne spread of infectious agents in the indoor environment. *Am. J. Infect. Control* **44**, S102–S108 (2016). doi: 10.1016/j.ajic.2016.06.003; pmid: 27590694
82. S. Niazi, R. Groth, K. Spann, G. R. Johnson, The role of respiratory droplet physicochemistry in limiting and promoting the airborne transmission of human coronaviruses: A critical review. *Environ. Pollut.* **276**, 115767 (2021). doi: 10.1016/j.envpol.2020.115767; pmid: 33243541
83. E. P. Vejerano, L. C. Marr, Physico-chemical characteristics of evaporating respiratory fluid droplets. *J. R. Soc. Interface* **15**, 20170939 (2018). doi: 10.1098/rsif.2017.0939; pmid: 29491178
84. B. Patterson, R. Wood, Is cough really necessary for TB transmission? *Tuberculosis* **117**, 31–35 (2019). doi: 10.1016/j.tube.2019.05.003; pmid: 31378265
85. H. Holmgren, E. Ljungström, A.-C. Almstrand, B. Bake, A.-C. Olin, Size distribution of exhaled particles in the range from 0.01 to 2.0 μm. *J. Aerosol Sci.* **41**, 439–446 (2010). doi: 10.1016/j.jaerosci.2010.02.011
86. L. Bourouiba, Turbulent gas clouds and respiratory pathogen emissions: Potential implications for reducing transmission of COVID-19. *JAMA* **323**, 1837–1838 (2020). doi: 10.1001/jama.2020.4756; pmid: 32215590
87. C. Kleinstreuer, Z. Zhang, Airflow and particle transport in the human respiratory system. *Annu. Rev. Fluid Mech.* **42**, 301–334 (2010). doi: 10.1146/annurev-fluid-121108-145453
88. G. R. Johnson, L. Morawska, The mechanism of breath aerosol formation. *J. Aerosol Med. Pulm. Drug Deliv.* **22**, 229–237 (2009). doi: 10.1089/jamp.2008.0720; pmid: 19415984
89. F. Koster, The experimental aerosol transmission of influenza virus. *Future Virol.* **8**, 969–981 (2013). doi: 10.2217/fvl.13.83
90. S. Asadi *et al*., Aerosol emission and superemission during human speech increase with voice loudness. *Sci. Rep.* **9**, 2348 (2019). doi: 10.1038/s41598-019-38808-z; pmid: 30787335
91. R. S. Papineni, F. S. Rosenthal, The size distribution of droplets in the exhaled breath of healthy human subjects. *J. Aerosol Med.* **10**, 105–116 (1997). doi: 10.1089/jam.1997.10.105; pmid: 10168531
92. M. W. Jennison, "Atomizing of Mouth and Nose Secretions into the Air as Revealed by High-Speed Photography" in *Aerobiology* (American Association for the Advancement of Science, ed. 17, 1942), pp. 106–128.
93. S. Shao *et al*., Risk assessment of airborne transmission of COVID-19 by asymptomatic individuals under different practical settings. *J. Aerosol Sci.* **151**, 105661 (2021). doi: 10.1016/j.jaerosci.2020.105661; pmid: 32968325
94. D. A. Edwards *et al*., Exhaled aerosol increases with COVID-19 infection, age, and obesity. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2021830118 (2021). doi: 10.1073/pnas.2021830118; pmid: 33563754
95. W. G. Lindsley *et al*., Quantity and size distribution of cough-generated aerosol particles produced by influenza patients during and after illness. *J. Occup. Environ. Hyg.* **9**, 443–449 (2012). doi: 10.1080/15459624.2012.684582; pmid: 22651099
96. M. Riediker, L. Morawska, Low exhaled breath droplet formation may explain why children are poor SARS-CoV-2 transmitters. *Aerosol Air Qual. Res.* **20**, 1513–1515 (2020). doi: 10.4209/aaqr.2020.06.0304
97. V. Stadnytskyi, C. E. Bax, A. Bax, P. Anfinrud, The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11875–11877 (2020). doi: 10.1073/pnas.2006874117; pmid: 32404416
98. M. Pan *et al*., Collection of viable aerosolized influenza virus and other respiratory viruses in a student health care center through water-based condensation growth. *MSphere* **2**, e00251-17 (2017). doi: 10.1128/mSphere.00251-17; pmid: 29034325
99. J. Y. Choi *et al*., Aerosol sampling in a hospital emergency room setting: A complementary surveillance method for the detection of respiratory viruses. *Front. Public Health* **6**, 174 (2018). doi: 10.3389/fpubh.2018.00174; pmid: 29963543
100. V. Knight, Viruses as agents of airborne contagion. *Ann. N. Y. Acad. Sci.* **353**, 147–156 (1980). doi: 10.1111/j.1749-6632.1980.tb18917.x; pmid: 6261640
101. W. G. Lindsley *et al*., Viable influenza A virus in airborne particles from human coughs. *J. Occup. Environ. Hyg.* **12**, 107–113 (2015). doi: 10.1080/15459624.2014.973113; pmid: 25523206
102. W. G. Lindsley *et al*., Distribution of airborne influenza virus and respiratory syncytial virus in an urgent care medical clinic. *Clin. Infect. Dis.* **50**, 693–698 (2010). doi: 10.1086/650457; pmid: 20100093
103. A. Totura *et al*., Small particle aerosol exposure of african green monkeys to MERS-CoV as a model for highly pathogenic coronavirus infection. *Emerg. Infect. Dis.* **26**, 2835–2843 (2020). doi: 10.3201/eid2612.201664; pmid: 32744989
104. D. K. Milton, M. P. Fabian, B. J. Cowling, M. L. Grantham, J. J. McDevitt, Influenza virus aerosols in human exhaled breath: Particle size, culturability, and effect of surgical masks. *PLOS Pathog.* **9**, e1003205 (2013). doi: 10.1371/journal.ppat.1003205; pmid: 23505369
105. W. G. Lindsley *et al*., Measurements of airborne influenza virus in aerosol particles from human coughs. *PLOS ONE* **5**, e15100 (2010). doi: 10.1371/journal.pone.0015100; pmid: 21152051
106. W. Yang, S. Elankumaran, L. C. Marr, Concentrations and size distributions of airborne influenza A viruses measured indoors at a health centre, a day-care centre and on aeroplanes. *J. R. Soc. Interface* **8**, 1176–1184 (2011). doi: 10.1098/rsif.2010.0686; pmid: 21300628
107. J. Ma *et al*., Coronavirus disease 2019 patients in earlier stages exhaled millions of severe acute respiratory syndrome coronavirus 2 per hour. *Clin. Infect. Dis.* **72**, e652–e654 (2021). doi: 10.1093/cid/ciaa1283; pmid: 32857833
108. P. Z. Chen *et al*., Heterogeneity in transmissibility and shedding SARS-CoV-2 via droplets and aerosols. *eLife* **10**, e65774 (2021). doi: 10.7554/eLife.65774; pmid: 33861198
109. S. B. Kwon *et al*., Study on the initial velocity distribution of exhaled air from coughing and speaking. *Chemosphere* **87**, 1260–1264 (2012). doi: 10.1016/j.chemosphere.2012.01.032; pmid: 22342283
110. J. D. Smith, C. D. Cappa, W. S. Drisdell, R. C. Cohen, R. J. Saykally, Raman thermometry measurements of free evaporation from liquid water droplets. *J. Am. Chem. Soc.* **128**, 12892–12898 (2006). doi: 10.1021/ja063579v; pmid: 17002384
111. L. Liu, J. Wei, Y. Li, A. Ooi, Evaporation and dispersion of respiratory droplets from coughing. *Indoor Air* **27**, 179–190 (2017). doi: 10.1111/ina.12297; pmid: 26945674
112. H. Wei *et al*., Aerosol microdroplets exhibit a stable pH gradient. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7272–7277 (2018). doi: 10.1073/pnas.1720488115; pmid: 29941550
113. K. Lin, L. C. Marr, Humidity-dependent decay of viruses, but not bacteria, in aerosols and droplets follows disinfection kinetics. *Environ. Sci. Technol.* **54**, 1024–1032 (2020). doi: 10.1021/acs.est.9b04959; pmid: 31886650
114. W. Yang, L. C. Marr, Dynamics of airborne influenza A viruses indoors and dependence on humidity. *PLOS ONE* **6**, e21481 (2011). doi: 10.1371/journal.pone.0021481; pmid: 21731764
115. J. Sznitman, Respiratory microflows in the pulmonary acinus. *J. Biomech.* **46**, 284–298 (2013). doi: 10.1016/j.jbiomech.2012.10.028; pmid: 23178038
116. W. C. Hinds, *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles* (Wiley, ed. 2, 1999).
117. D. Parienta *et al*., Theoretical analysis of the motion and evaporation of exhaled respiratory droplets of mixed composition. *J. Aerosol Sci.* **42**, 1–10 (2011). doi: 10.1016/j.jaerosci.2010.10.005
118. R. Mittal, R. Ni, J.-H. Seo, The flow physics of COVID-19. *J. Fluid Mech.* **894**, F2 (2020). doi: 10.1017/jfm.2020.330
119. V. Vuorinen *et al*., Modelling aerosol transport and virus exposure with numerical simulations in relation to SARS-CoV-2 transmission by inhalation indoors. *Saf. Sci.* **130**, 104866 (2020). doi: 10.1016/j.ssci.2020.104866; pmid: 32834511
120. X. Xie, Y. Li, A. T. Y. Chwang, P. L. Ho, W. H. Seto, How far droplets can move in indoor environments—Revisiting the Wells evaporation-falling curve. *Indoor Air* **17**, 211–225 (2007). doi: 10.1111/j.1600-0668.2007.00469.x; pmid: 17542834
121. L. Bourouiba, E. Dehandschoewercker, J. W. M. Bush, Violent expiratory events: On coughing and sneezing. *J. Fluid Mech.* **745**, 537–563 (2014). doi: 10.1017/jfm.2014.88
122. J. A. Posada, J. Redrow, I. Celik, A mathematical model for predicting the viability of airborne viruses. *J. Virol. Methods* **164**, 88–95 (2010). doi: 10.1016/j.jviromet.2009.12.004; pmid: 20025904
123. P. Dabisch *et al*., The influence of temperature, humidity, and simulated sunlight on the infectivity of SARS-CoV-2 in aerosols. *Aerosol Sci. Technol.* **55**, 142–153 (2021). doi: 10.1080/02786826.2020.1829536
124. E. C. Pirtle, G. W. Beran, Virus survival in the environment. *Rev. Sci. Tech.* **10**, 733–748 (1991). doi: 10.20506/rst.10.3.570; pmid: 1782426
125. J. W. Tang, The effect of environmental parameters on the survival of airborne infectious agents. *J. R. Soc. Interface* **6**, S737–S746 (2009). doi: 10.1098/rsif.2009.0227.focus; pmid: 19773291
126. D. Welch *et al*., Far-UVC light: A new tool to control the spread of airborne-mediated microbial diseases. *Sci. Rep.* **8**, 2752 (2018). doi: 10.1038/s41598-018-21058-w; pmid: 29426899
127. J. J. McDevitt, S. N. Rudnick, L. J. Radonovich, Aerosol susceptibility of influenza virus to UV-C light. *Appl. Environ.*

Microbiol. **78**, 1666–1669 (2012). doi: 10.1128/AEM.06960-11; pmid: 22226954
128. K. Lin, D. Yee-Tak Fong, B. Zhu, J. Karlberg, Environmental factors on the SARS epidemic: Air temperature, passage of time and multiplicative effect of hospital infection. *Epidemiol. Infect.* **134**, 223–230 (2006). doi: 10.1017/S0950268805005054; pmid: 16490124
129. L. C. Marr, J. W. Tang, J. Van Mullekom, S. S. Lakdawala, Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. *J. R. Soc. Interface* **16**, 20180298 (2019). doi: 10.1098/rsif.2018.0298; pmid: 30958176
130. E. R. McFadden Jr *et al*., Thermal mapping of the airways in humans. *J. Appl. Physiol.* **58**, 564–570 (1985). doi: 10.1152/jappl.1985.58.2.564; pmid: 3980358
131. D. A. Tyrrell, R. Parsons, Some virus isolations from common colds. III. Cytopathic effects in tissue cultures. *Lancet* **275**, 239–242 (1960). doi: 10.1016/S0140-6736(60)90168-9; pmid: 13840115
132. K. H. Chan *et al*., The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Adv. Virol.* **2011**, 734690 (2011). doi: 10.1155/2011/734690; pmid: 22312351
133. A. W. H. Chin *et al*., Stability of SARS-CoV-2 in different environmental conditions. *Lancet Microbe* **1**, e10 (2020). doi: 10.1016/S2666-5247(20)30003-3; pmid: 32835322
134. M. Moriyama, W. J. Hugentobler, A. Iwasaki, Seasonality of respiratory viral infections. *Annu. Rev. Virol.* **7**, 83–101 (2020). doi: 10.1146/annurev-virology-012420-022445; pmid: 32196426
135. A. C. Lowen, J. Steel, S. Mubareka, P. Palese, High temperature (30°C) blocks aerosol but not contact transmission of influenza virus. *J. Virol.* **82**, 5650–5652 (2008). doi: 10.1128/JVI.00325-08; pmid: 18367530
136. J. S. Walker *et al*., Accurate representations of the microphysical processes occurring during the transport of exhaled aerosols and droplets. *ACS Cent. Sci.* **7**, 200–209 (2021). doi: 10.1021/acscentsci.0c01522; pmid: 33532579
137. K. M. Gustin *et al*., Environmental conditions affect exhalation of H3N2 seasonal and variant influenza viruses and respiratory droplet transmission in ferrets. *PLOS ONE* **10**, e0125874 (2015). doi: 10.1371/journal.pone.0125874; pmid: 25969995
138. K. A. Kormuth *et al*., Influenza virus infectivity is retained in aerosols and droplets independent of relative humidity. *J. Infect. Dis.* **218**, 739–747 (2018). doi: 10.1093/infdis/jiy221; pmid: 29878137
139. J. R. Songer, Influence of relative humidity on the survival of some airborne viruses. *Appl. Microbiol.* **15**, 35–42 (1967). doi: 10.1128/am.15.1.35-42.1967; pmid: 4291670
140. M. Schuit *et al*., The influence of simulated sunlight on the inactivation of influenza virus in aerosols. *J. Infect. Dis.* **221**, 372–378 (2020). doi: 10.1093/infdis/jiz582; pmid: 31778532
141. M. Buonanno, D. Welch, I. Shuryak, D. J. Brenner, Far-UVC light (222 nm) efficiently and safely inactivates airborne human coronaviruses. *Sci. Rep.* **10**, 10285 (2020). doi: 10.1038/s41598-020-67211-2; pmid: 32581288
142. C. S. Heilingloh *et al*., Susceptibility of SARS-CoV-2 to UV irradiation. *Am. J. Infect. Control* **48**, 1273–1275 (2020). doi: 10.1016/j.ajic.2020.07.031; pmid: 32763344
143. Y. Ye, P. H. Chang, J. Hartert, K. R. Wigginton, Reactivity of enveloped virus genome, proteins, and lipids with free chlorine and UV_{254}. *Environ. Sci. Technol.* **52**, 7698–7708 (2018). doi: 10.1021/acs.est.8b00824; pmid: 29886734
144. Y. Li *et al*., Role of ventilation in airborne transmission of infectious agents in the built environment - a multidisciplinary systematic review. *Indoor Air* **17**, 2–18 (2007). doi: 10.1111/j.1600-0668.2006.00445.x; pmid: 17257148
145. J. W. Tang, Y. Li, I. Eames, P. K. S. Chan, G. L. Ridgway, Factors involved in the aerosol transmission of infection and control of ventilation in healthcare premises. *J. Hosp. Infect.* **64**, 100–114 (2006). doi: 10.1016/j.jhin.2006.05.022; pmid: 16916564
146. C.-R. Du *et al*., Effect of ventilation improvement during a tuberculosis outbreak in underventilated university buildings. *Indoor Air* **30**, 422–432 (2020). doi: 10.1111/ina.12639; pmid: 31883403
147. S. E. Hwang, J. H. Chang, B. Oh, J. Heo, Possible aerosol transmission of COVID-19 associated with an outbreak in an apartment in Seoul, South Korea, 2020. *Int. J. Infect. Dis.* **104**, 73–76 (2021). doi: 10.1016/j.ijid.2020.12.035; pmid: 33346125
148. H. Qian, X. Zheng, Ventilation control for airborne transmission of human exhaled bio-aerosols in buildings. *J. Thorac. Dis.* **10**, S2295–S2304 (2018). doi: 10.21037/jtd.2018.01.24; pmid: 30116608
149. World Health Organization (WHO), "Roadmap to improve and ensure good indoor ventilation in the context of COVID-19" (2021); www.who.int/publications/i/item/9789240021280.
150. W. G. Lindsley *et al*., Efficacy of portable air cleaners and masking for reducing indoor exposure to simulated exhaled SARS-CoV-2 aerosols — United States, 2021. *MMWR Morb. Mortal. Wkly. Rep.* **70**, 972–976 (2021). doi: 10.15585/mmwr.mm7027e1; pmid: 34237047
151. S. R. Narayanan, S. Yang, Airborne transmission of virus-laden aerosols inside a music classroom: Effects of portable purifiers and aerosol injection rates. *Phys. Fluids* **33**, 033307 (2021). doi: 10.1063/5.0042474; pmid: 33746493
152. J. Curtius, M. Granzin, J. Schrod, Testing mobile air purifiers in a school classroom: Reducing the airborne transmission risk for SARS-CoV-2. *Aerosol Sci. Technol.* **55**, 586–599 (2021). doi: 10.1080/02786826.2021.1877257
153. J. Lessler *et al*., Household COVID-19 risk and in-person schooling. *Science* **372**, 1092–1097 (2021). doi: 10.1126/science.abh2939; pmid: 33927057
154. C. Darquenne, Aerosol deposition in health and disease. *J. Aerosol Med. Pulm. Drug Deliv.* **25**, 140–147 (2012). doi: 10.1089/jamp.2011.0916; pmid: 22686623
155. C. Darquenne, Deposition Mechanisms. *J. Aerosol Med. Pulm. Drug Deliv.* **33**, 181–185 (2020). doi: 10.1089/jamp.2020.29029.cd; pmid: 32598200
156. A. E. Haddrell *et al*., Pulmonary aerosol delivery and the importance of growth dynamics. *Ther. Deliv.* **8**, 1051–1061 (2017). doi: 10.4155/tde-2017-0093; pmid: 29125064
157. S. Guha, P. Hariharan, M. R. Myers, Enhancement of ICRP's lung deposition model for pathogenic bioaerosols. *Aerosol Sci. Technol.* **48**, 1226–1235 (2014). doi: 10.1080/02786826.2014.975334
158. D. K. Milton, A rosetta stone for understanding infectious drops and aerosols. *J. Pediatric Infect. Dis. Soc.* **9**, 413–415 (2020). doi: 10.1093/jpids/piaa079; pmid: 32706376
159. P. Hofemeier, K. Koshiyama, S. Wada, J. Sznitman, One (sub-)acinus for all: Fate of inhaled aerosols in heterogeneous pulmonary acinar structures. *Eur. J. Pharm. Sci.* **113**, 53–63 (2018). doi: 10.1016/j.ejps.2017.09.033; pmid: 28954217
160. L. Zhang, Z. Gu, C. Yu, Y. Zhang, Y. Cheng, Surface charges on aerosol particles – accelerating particle growth rate and atmospheric pollution. *Indoor Built Environ.* **25**, 437–440 (2016). doi: 10.1177/1420326X16643799
161. M. S. P. Islam *et al*., A review of respiratory anatomical development, air flow characterization and particle deposition. *Int. J. Environ. Res. Public Health* **17**, 380 (2020). doi: 10.3390/ijerph17020380; pmid: 31935991
162. J. K. Mutuku, W.-C. Hou, W.-H. Chen, Two-phase flow dynamics and $PM_{2.5}$ deposition in healthy and obstructed human airways during inhalation. *Aerosol Air Qual. Res.* **20**, 1091–1110 (2020). doi: 10.4209/aaqr.2020.03.0107
163. W.-H. Chen, K.-H. Lee, J. K. Mutuku, C.-J. Hwang, Flow dynamics and $PM_{2.5}$ deposition in healthy and asthmatic airways at different inhalation statuses. *Aerosol Air Qual. Res.* **18**, 866–883 (2018). doi: 10.4209/aaqr.2018.02.0058
164. E. K. Alidjinou *et al*., Spatial and temporal virus load dynamics of SARS-CoV-2: A single-center cohort study. *Diagnostics* **11**, 427 (2021). doi: 10.3390/diagnostics11030427; pmid: 33802451
165. A. Weiss, M. Jellingsø, M. O. A. Sommer, Spatial and temporal dynamics of SARS-CoV-2 in COVID-19 patients: A systematic review and meta-analysis. *EBioMedicine* **58**, 102916 (2020). doi: 10.1016/j.ebiom.2020.102916; pmid: 32711256
166. N. Eichler *et al*., Transmission of severe acute respiratory syndrome coronavirus 2 during border quarantine and air travel, New Zealand (Aotearoa). *Emerg. Infect. Dis.* **27**, 1274–1278 (2021). doi: 10.3201/eid2705.210514; pmid: 33734063
167. M. M. Arons *et al*., Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N. Engl. J. Med.* **382**, 2081–2090 (2020). doi: 10.1056/NEJMoa2008457; pmid: 32329971
168. 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
169. D. P. Oran, E. J. Topol, Prevalence of asymptomatic SARS-CoV-2 infection: A Narrative Review. *Ann. Intern. Med.* **173**, 362–367 (2020). doi: 10.7326/M20-3012; pmid: 32491919
170. X. He *et al*., Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020). doi: 10.1038/s41591-020-0869-5; pmid: 32296168
171. M. Galanti *et al*., Rates of asymptomatic respiratory virus infection across age groups. *Epidemiol. Infect.* **147**, e176 (2019). doi: 10.1017/S0950268819000505; pmid: 31063096
172. N. H. L. Leung, C. Xu, D. K. M. Ip, B. J. Cowling, Review article: The fraction of influenza virus infections that are asymptomatic: A systematic review and meta-analysis. *Epidemiology* **26**, 862–872 (2015). doi: 10.1097/EDE.0000000000000340; pmid: 26133025
173. F. Carrat *et al*., Time lines of infection and disease in human influenza: A review of volunteer challenge studies. *Am. J. Epidemiol.* **167**, 775–785 (2008). doi: 10.1093/aje/kwm375; pmid: 18230677
174. L. Zou *et al*., SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N. Engl. J. Med.* **382**, 1177–1179 (2020). doi: 10.1056/NEJMc2001737; pmid: 32074444
175. K. A. Walsh *et al*., SARS-CoV-2 detection, viral load and infectivity over the course of an infection. *J. Infect.* **81**, 357–371 (2020). doi: 10.1016/j.jinf.2020.06.067; pmid: 32615199
176. S. E. Eikenberry *et al*., To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infect. Dis. Model.* **5**, 293–308 (2020). doi: 10.1016/j.idm.2020.04.001; pmid: 32355904
177. C. J. Worby, H.-H. Chang, Face mask use in the general population and optimal resource allocation during the COVID-19 pandemic. *Nat. Commun.* **11**, 4049 (2020). doi: 10.1038/s41467-020-17922-x; pmid: 32792562
178. 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
179. M. Gandhi, L. C. Marr, Uniting infectious disease and physical science principles on the importance of face masks for COVID-19. *Med* **2**, 29–32 (2021). doi: 10.1016/j.medj.2020.12.008; pmid: 33521753
180. F. Drewnick *et al*., Aerosol filtration efficiency of household materials for homemade face masks: Influence of material properties, particle size, particle electrical charge, face velocity, and leaks. *Aerosol Sci. Technol.* **55**, 63–79 (2021). doi: 10.1080/02786826.2020.1817846
181. H. Whiley, T. P. Keerthirathne, M. A. Nisar, M. A. F. White, K. E. Ross, Viral filtration efficiency of fabric masks compared with surgical and N95 masks. *Pathogens* **9**, 762 (2020). doi: 10.3390/pathogens9090762; pmid: 32957638
182. H. Ueki *et al*., Effectiveness of face masks in preventing airborne transmission of SARS-CoV-2. *mSphere* **5**, e00637-20 (2020). doi: 10.1128/mSphere.00637-20; pmid: 33087517
183. M. Klompas *et al*., Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) From Asymptomatic and Presymptomatic Individuals in Healthcare Settings Despite Medical Masks and Eye Protection. *Clin. Infect. Dis.* 10.1093/cid/ciab218 (2021). doi: 10.1093/cid/ciab218; pmid: 33704451
184. L. Goldberg *et al*., SARS-CoV-2 infection among health care workers despite the use of surgical masks and physical distancing—the role of airborne transmission. *Open Forum Infect. Dis.* **8**, ofab036 (2021). doi: 10.1093/ofid/ofab036; pmid: 33732749
185. M. Klompas *et al*., A SARS-CoV-2 cluster in an acute care hospital. *Ann. Intern. Med.* **174**, 794–802 (2021). doi: 10.7326/M20-7567; pmid: 33556277
186. T. C. Bulfone, M. Malekinejad, G. W. Rutherford, N. Razani, Outdoor transmission of SARS-CoV-2 and other respiratory viruses: A systematic review. *J. Infect. Dis.* **223**, 550–561 (2021). doi: 10.1093/infdis/jiaa742; pmid: 33249484
187. N. R. Faria *et al*., Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815–821 (2021). doi: 10.1126/science.abh2644; pmid: 33853970
188. N. G. Davies *et al*., Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021). doi: 10.1126/science.abg3055; pmid: 33658326
189. M. Kang *et al*., Probable evidence of fecal aerosol transmission of SARS-CoV-2 in a high-rise building. *Ann. Intern. Med.* **173**, 974–980 (2020). doi: 10.7326/M20-0928; pmid: 32870707

190. G. A. Somsen, C. van Rijn, S. Kooij, R. A. Bem, D. Bonn, Small droplet aerosols in poorly ventilated spaces and SARS-CoV-2 transmission. *Lancet Respir. Med.* **8**, 658–659 (2020). doi: 10.1016/S2213-2600(20)30245-9; pmid: 32473123
191. L. Morawska *et al*., How can airborne transmission of COVID-19 indoors be minimised? *Environ. Int.* **142**, 105832 (2020). doi: 10.1016/j.envint.2020.105832; pmid: 32521345
192. S. N. Rudnick, D. K. Milton, Risk of indoor airborne infection transmission estimated from carbon dioxide concentration. *Indoor Air* **13**, 237–245 (2003). doi: 10.1034/j.1600-0668.2003.00189.x; pmid: 12950586
193. Z. Peng, J. L. Jimenez, Exhaled CO_2 as COVID-19 infection risk proxy for different indoor environments and activities. *Environ. Sci. Technol. Lett.* **8**, 392–397 (2021). doi: 10.1021/acs.estlett.1c00183
194. F. Villanueva *et al*., Assessment of CO_2 and aerosol ($PM_{2.5}$, PM_{10}, UFP) concentrations during the reopening of schools in the COVID-19 pandemic: The case of a metropolitan area in Central-Southern Spain. *Environ. Res.* **197**, 111092 (2021). doi: 10.1016/j.envres.2021.111092; pmid: 33785326
195. C. Chen *et al*., The effectiveness of an air cleaner in controlling droplet/aerosol particle dispersion emitted from a patient's mouth in the indoor environment of dental clinics. *J. R. Soc. Interface* **7**, 1105–1118 (2010). doi: 10.1098/rsif.2009.0516; pmid: 20031985
196. M. Z. Bazant, J. W. M. Bush, A guideline to limit indoor airborne transmission of COVID-19. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2018995118 (2021). doi: 10.1073/pnas.2018995118; pmid: 33858987
197. E. Petersen *et al*., Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* **20**, e238–e244 (2020). doi: 10.1016/S1473-3099(20)30484-9; pmid: 32628905
198. X.-Y. Hao, Q. Lv, F. D. Li, Y. F. Xu, H. Gao, The characteristics of hDPP4 transgenic mice subjected to aerosol MERS coronavirus infection via an animal nose-only exposure device. *Animal Model Exp. Med.* **2**, 269–281 (2019). doi: 10.1002/ame2.12088; pmid: 31942559
199. S. L. Bixler *et al*., Aerosol Exposure of Cynomolgus Macaques to SARS-CoV-2 Results in More Severe Pathology than Existing Models. bioRxiv 2021.04.27.441510 [Preprint] (2021). doi: 10.1101/2021.04.27.441510
200. S. S. Lakdawala *et al*., Eurasian-origin gene segments contribute to the transmissibility, aerosol release, and morphology of the 2009 pandemic H1N1 influenza virus. *PLOS Pathog.* **7**, e1002443 (2011). doi: 10.1371/journal.ppat.1002443; pmid: 22241979
201. J. Zhou *et al*., Defining the sizes of airborne particles that mediate influenza transmission in ferrets. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2386–E2392 (2018). doi: 10.1073/pnas.1716771115; pmid: 29463703
202. J. S. Nguyen-Van-Tam *et al*., Minimal transmission in an influenza A (H3N2) human challenge-transmission model within a controlled exposure environment. *PLOS Pathog.* **16**, e1008704 (2020). doi: 10.1371/journal.ppat.1008704; pmid: 32658939
203. K. A. Kormuth *et al*., Environmental persistence of influenza viruses is dependent upon virus type and host origin. *MSphere* **4**, e00552-19 (2019). doi: 10.1128/mSphere.00552-19; pmid: 31434749
204. P. J. Bueno de Mesquita, C. J. Noakes, D. K. Milton, Quantitative aerobiologic analysis of an influenza human challenge-transmission trial. *Indoor Air* **30**, 1189–1198 (2020). doi: 10.1111/ina.12701; pmid: 32542890
205. N. H. L. Leung, Transmissibility and transmission of respiratory viruses. *Nat. Rev. Microbiol.* **19**, 528–545 (2021). doi: 10.1038/s41579-021-00535-6; pmid: 33753932
206. F. M. Guerra *et al*., The basic reproduction number (R_0) of measles: A systematic review. *Lancet Infect. Dis.* **17**, e420–e428 (2017). doi: 10.1016/S1473-3099(17)30307-9; pmid: 28757186

ACKNOWLEDGMENTS

C.C.W. thanks D. M. Neumark, K. Liu, I. Gonda, and Y.-Y. Cheng for helpful discussions. **Funding:** C.C.W. is supported by the Ministry of Science and Technology (MOST 109-2113-M-110-011 and MOST 109-2621-110-006) and the Higher Education Sprout Project of the Ministry of Education, Taiwan, ROC. K.A.P. is supported by the US NSF Center for Aerosol Impacts on Chemistry of the Environment, USA. J.L.J. is supported by the US National Science Foundation (AGS-1822664). L.C.M. is supported by the National Institute of Allergy and Infectious Diseases Center of Excellence in Influenza Research and Surveillance (HHSN272201400007C) and the NSF National Nanotechnology Coordinated Infrastructure (ECCS 1542100 and ECCS 2025151). **Competing interests:** L.C.M. has served on Advisory Boards for Crossfit and Phylagen, has served as a paid consultant for The MITRE Corporation and Smiths Detection, and was a paid reviewer for the Alfred P. Sloan Foundation. She is an unpaid member of the National Academies of Sciences, Engineering, and Medicine Board on Environmental Studies and Toxicology and the Committee on Public Health Interventions and Countermeasures for Advancing Pandemic and Seasonal Influenza Preparedness and Response. The authors declare no other competing interests.

This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

10.1126/science.abd9149

RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

Prenatal maternal infection promotes tissue-specific immunity and inflammation in offspring

Ai Ing Lim, Taryn McFadden, Verena M. Link, Seong-Ji Han, Rose-Marie Karlsson, Apollo Stacy, Taylor K. Farley, Djalma S. Lima-Junior, Oliver J. Harrison, Jigar V. Desai, Michail S. Lionakis, Han-Yu Shih, Heather A. Cameron, Yasmine Belkaid*

INTRODUCTION: One fundamental property of the immune system is its ability to develop memory of previous encounters, resulting in enhanced responsiveness to subsequent challenges. This phenomenon includes not only the adaptive immune system but also innate cells and tissue stem cells. The concept of host imprinting by infection leading to altered responses to subsequent challenge is of particular interest in the context of pregnancy, which represents a fundamental developmental window for the immune system.

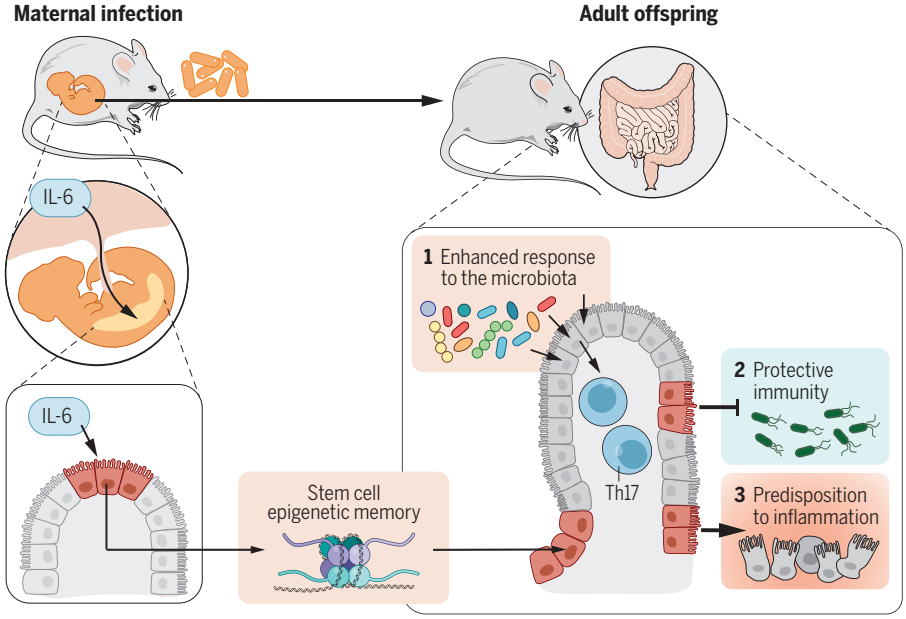
RATIONALE: Most infections encountered by mammals, including those experienced during pregnancy, are mild and transient. How these infections shape offspring tissue immunity and tissue predisposition to inflammatory disorders in the long term remains to be addressed.

RESULTS: Infection of timed-pregnant dams (day 10.5) with an attenuated strain of the food-borne pathogen *Yersinia pseudotuberculosis* (*yopM*) was transient and maternally restricted. Adult offspring of previously infected dams harbored a higher number of T helper 17 (T_H17) cells but no other cell subsets in the small and large intestinal lamina propria. No changes were observed at other barrier tissues. Transfer of transgenic T cells specific for a commensal antigen revealed that maternal infection affected the offspring intestinal milieu in a manner that enhanced T_H17 cell reactivity toward the microbiota. Cross-fostering experiments demonstrated that the increased T_H17 cells resulting from maternal infection was imprinted in utero. Among various inflammatory mediators, the cytokine interleukin-6 (IL-6) was significantly increased in the serum of dams infected with *yopM*. Injection of IL-6 alone to pregnant dams significantly increased T_H17 cell numbers within the guts of offspring. Conversely, blockade of IL-6 in infected dams prevented this increase in offspring. Furthermore, injection of IL-6 to germ-free pregnant dams and conventionalization of their offspring after weaning revealed that prenatal establishment and postnatal maintenance of IL-6–mediated tissue imprinting are independent of the maternal microbiota but allow the offspring to mount enhanced T_H17 cell responses to postnatal microbiota exposure.

IL6RA was expressed in all fetal intestinal epithelial cells, and specific deletion of IL6RA from epithelial cells significantly reduced Th17 responses within the gut of offspring from IL-6–exposed dams. Using complementary approaches, including assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) and single-cell RNA sequencing, we found that increased IL-6 during pregnancy had immediate effects (on fetal cells) and long-term effects (in the adult offspring) on epithelial stem cell chromatin accessibility and downstream epithelial cell function, particularly increased expression of genes encoding for antigen presentation machinery and antimicrobial peptides.

Altered epithelial activation status suggested the possibility that this phenomenon may be associated with enhanced antimicrobial defense. To address this, we used an acute model of oral infection with *Salmonella* Typhimurium. Offspring from dams previously infected with *yopM* or injected with IL-6 during pregnancy developed enhanced resistance to *Salmonella* oral infection and dissemination. No differences were observed in controlling cutaneous *Candida albicans* infection, further supporting the idea that maternal imprinting of offspring is restricted to the gut compartment. However, enhanced exposure to IL-6 during pregnancy was associated with enhanced gut pathology in the context of naïve T cell transfer and dextran sulfate sodium–mediated colitis.

CONCLUSION: Our work proposes that a transient, mild infection encountered during prenatal development can impose lasting alterations to gut epithelial stem cells, resulting in an altered threshold of activation and enhanced resistance to gut infection. The impact of maternal infection was tissue specific and predominantly mediated by a single cytokine, IL-6, acting on epithelial stem cells during fetal development. Although this phenomenon can be coopted by the fetus to develop optimal immune fitness, altered offspring immunity imposed by maternal infection comes at the cost of enhanced susceptibility to mucosal inflammation. ■



Maternal infection promotes offspring intestine-specific immunity and inflammation. The direct response of fetal intestinal epithelial cells to IL-6 during maternal infection confers an enduring epigenetic memory to adult intestinal epithelial stem cells. As such, offspring epithelial cells exhibit enhanced reactivity toward the microbiota and heightened ability to control oral infection. However, these responses come at the cost of greater predisposition to gut inflammation.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ybelkaid@niaid.nih.gov
Cite this article as A. I. Lim *et al.*, *Science* **373**, eabf3002 (2021). DOI: 10.1126/science.abf3002

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abf3002

RESEARCH ARTICLE

IMMUNOLOGY

Prenatal maternal infection promotes tissue-specific immunity and inflammation in offspring

Ai Ing Lim[1], Taryn McFadden[1], Verena M. Link[1], Seong-Ji Han[1], Rose-Marie Karlsson[2], Apollo Stacy[1,3], Taylor K. Farley[1,4], Djalma S. Lima-Junior[1], Oliver J. Harrison[1]†, Jigar V. Desai[5], Michail S. Lionakis[5], Han-Yu Shih[6], Heather A. Cameron[2], Yasmine Belkaid[1,7]*

The immune system has evolved in the face of microbial exposure. How maternal infection experienced at distinct developmental stages shapes the offspring immune system remains poorly understood. Here, we show that during pregnancy, maternally restricted infection can have permanent and tissue-specific impacts on offspring immunity. Mechanistically, maternal interleukin-6 produced in response to infection can directly impose epigenetic changes on fetal intestinal epithelial stem cells, leading to long-lasting impacts on intestinal immune homeostasis. As a result, offspring of previously infected dams develop enhanced protective immunity to gut infection and increased inflammation in the context of colitis. Thus, maternal infection can be coopted by the fetus to promote long-term, tissue-specific fitness, a phenomenon that may come at the cost of predisposition to inflammatory disorders.

One fundamental property of the immune system is its ability to develop memory of previous encounters, resulting in enhanced responsiveness to subsequent challenges. Although immunological memory was initially characterized as a defining property of the adaptive immune system, a growing body of evidence supports the idea that components of the innate system can also be profoundly affected by inflammatory or infectious encounters, a phenomenon referred to as trained immunity (*1*). Tissue stem cells can also develop memory of previous insults, resulting in accelerated responses to subsequent challenges (*2*, *3*). Innate imprinting is usually transient, lasting a few weeks to months after exposure. However, parental experiences can be transferred across generations in invertebrates, a phenomenon that is also hypothesized to exist in vertebrates (*4*–*6*). Education of innate or stem cells by previous inflammatory insults has been associated with epigenetic reprograming (*2*, *3*, *7*), although the mechanisms underlying this fundamental property remain largely elusive and represent an active area of investigation. This revised understanding of immunity as a phenomenon encompassing the experiences of the whole organism as opposed to specialized cells provides a framework with which to understand the enhanced immune fitness observed in animals raised in physiological environments, in contrast to animals raised under highly controlled, laboratory settings (*8*, *9*).

The concept of host immune imprinting is of particular interest in the context of pregnancy, which represents a fundamental developmental window for the immune system. Pregnancy is associated with specific maternal adaptations that serve to maintain tolerance to the fetus (*10*). At the same time, the mother can contribute to offspring protection against pathogens through the transfer of immunomodulatory factors, including maternal antibodies and immune cells, in utero and through breastfeeding (*11*–*13*). Maternal immune activation has also been associated with altered offspring systemic T cell responses (*14*). However, how infection experienced by the mother during pregnancy shapes offspring tissue immunity in the long term remains poorly understood. Indeed, most of our current understanding derives from studies illustrating detrimental outcomes of severe maternal inflammation or infections with pathogens that are able to cross the placental barrier (*15*). How defined stressors encountered during pregnancy alter offspring predisposition to inflammatory disorders remains unclear. Mammals and their immune systems evolved in the face of a microbially rich environment, and frequent exposure to pathogens and associated protective immunity represents the norm rather than the exception. Host survival also requires prioritization of defense strategies that protect the host from tissue damage, in some cases by limiting infection severity without affecting the pathogen load, a phenomenon referred to as disease tolerance (*16*, *17*). As a result, by far most microbial encounters are expected to be associated with mild or no symptoms.

Here, we explored the possibility that maternal infection experienced during pregnancy could affect the offspring immune system in the long term. On the basis of the fundamental developmental processes that occur in utero, we proposed that specific signals experienced at specific developmental windows may imprint the fetus in a discrete and tissue-specific manner.

Maternally restricted infection induces tissue-specific alterations to offspring immunity

To assess the impact of maternally restricted infection on offspring immunity, we used an attenuated strain of the food-borne pathogen *Yersinia pseudotuberculosis*. This specific strain, which harbors a mutation in an effector protein (YopM) translocated by the type III secretion system, causes a mild and transient infection (*18*, *19*). Timed-pregnant dams were orally infected with the mutant strain (*yopM*) at gestation day 10.5 (Fig. 1A). This specific time point represents a critical developmental stage for the fetal immune system, characterized by the emergence of hematopoietic stem cells and the initiation of hematopoiesis (*20*, *21*). After *yopM* infection, dams experienced transient weight loss from days 5 to 8 after infection, but no other symptoms were observed (fig. S1A). Timed-pregnant dams transiently shed *yopM* in their stool (fig. S1B), and only 20% of pregnant females experienced transient bacterial dissemination to the liver and spleen (fig. S1C). In a manner comparable to nonpregnant mice (*18*), *yopM* was controlled in the intestines by day 5 after infection and cleared by day 9 (fig. S1B). No bacteria were detected in the fetal intestine (fig. S1B), placenta, or fetal liver (fig. S1C), as assessed by culture and polymerase chain reaction (PCR), at any time point tested. Moreover, infection did not affect litter sizes (fig. S1D). Thus, *yopM* infection represents a model of mild, maternally restricted infection.

We next compared the lymphocyte compartments of the adult offspring delivered by naïve versus previously infected dams. To this end, we characterized the number, frequency, and phenotype of T cells and innate lymphoid cells (ILCs) at barrier tissues, including the skin, the lung, and the small (siLP) and large intestinal lamina propria (liLP) of 5- to 8-week-old offspring (fig. S1E). Maternal infection during

[1]Metaorganism Immunity Section, Laboratory of Immune System Biology and Laboratory of Host Immunity and Microbiome, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [2]Section on Neuroplasticity, Mood and Anxiety Disorders Program, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA. [3]Postdoctoral Research Associate Training Program, National Institute of General Medical Sciences, Bethesda, MD 20892, USA. [4]Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7FY, UK. [5]Fungal Pathogenesis Section, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [6]Neuro-Immune Regulome Unit, National Eye Institute, National Institutes of Health, Bethesda, MD 20892, USA. [7]NIAID Microbiome Program, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.
*Corresponding author. Email: ybelkaid@niaid.nih.gov
†Present address: Center for Fundamental Immunology, Benaroya Research Institute, Seattle, WA 98101, USA.

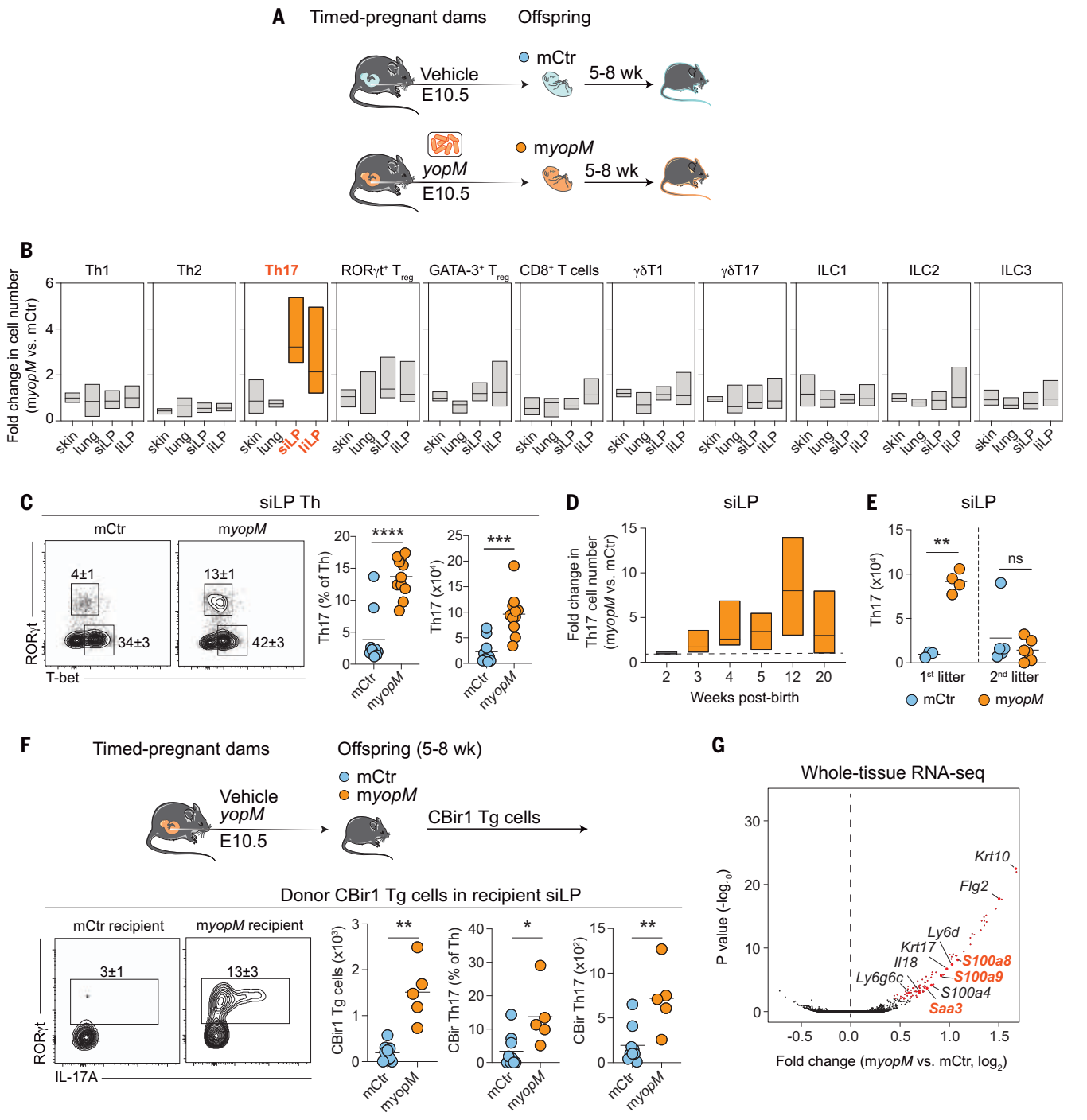


Fig. 1. Maternally restricted infection induces tissue-specific alterations to offspring immunity. (**A**) Timed-pregnant dams were orally administered vehicle (PBS) or *Y. pseudotuberculosis yopM* on gestation day 10.5 (E10.5), and their offspring, mCtr or m*yopM*, were analyzed at 5 to 8 weeks of age. (**B**) Fold change in cell number of lymphocyte subsets in ear skin, lung, siLP, and liLP of m*yopM* offspring compared with the mean of mCtr offspring. Cells were gated according to the schematic provided in fig. S1E. (**C**) Left: Representative contour plots of transcription factors expressed by offspring siLP T_H subsets (gated on live $CD45^+CD90.2^+TCR\beta^+CD4^+Foxp3^-$). Right: Frequency of T_H17 ($ROR\gamma t^+$) cells among total T_H cells, and total siLP T_H17 cell number. (**D**) Fold change in T_H17 cell number in the siLP of m*yopM* offspring compared with the mean of mCtr offspring at corresponding ages. (**E**) Total siLP T_H17 cell number in the offspring of vehicle-treated dams or dams infected during pregnancy (1st) or the offspring from a subsequent pregnancy of the same dams (2nd). (**F**) Donor CBir1 Tg cells (CD45.1) were transferred to congenic mCtr or m*yopM* offspring (CD45.2). Donor cells in the siLP of recipient offspring were analyzed for transcription factor expression and cytokine production after restimulation at days 10 to 12 after transfer. Contour plots were gated on live $CD45.1^+CD90.2^+TCR\beta^+CD4^+Foxp3^-$. (**G**) Whole-tissue RNA-seq of ileum from m*yopM* offspring compared with mCtr offspring. Up-regulated genes are denoted in red. (A to G) Data are representative of three independent experiments, each experiment with one or two pregnant dams per group using three to five offspring per pregnancy. Each dot represents an individual mouse. Numbers in representative flow plots indicate mean ± SD. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; ns, not significant (two-tailed unpaired Student's *t* test). See also fig. S1.

pregnancy was associated with a significant and selective increase in the number and frequency of RORγt-expressing $CD4^+$ T (T_H17) cells in the siLP of offspring compared with animals delivered from naïve dams (Fig. 1, B and C). The increase in T_H17 cell number was restricted to the gut compartment and was not detected at other assessed barrier sites (Fig. 1B). Furthermore, no differences were observed in the number of other lymphocyte subsets, including T-bet–expressing $CD4^+$ T (T_H1) cells, GATA3-expressing $CD4^+$ T (T_H2) cells, Foxp3-expressing $CD4^+$ regulatory T cells [RORgt$^+$ regulatory T cell (T_{reg}) and $GATA3^+$ T_{reg}], $CD8^+$ T cells, γδ T cells (γδT1 and γδT17), and ILCs (ILC1, ILC2, and ILC3) in all the compartments assessed (Fig. 1B and fig. S1F). Increased accumulation of T_H17 cells within the siLP was detectable after weaning (at 3 weeks) and persisted throughout adulthood (up to 20 weeks) (Fig. 1D and fig. S1G). The impact of infection on the offspring gut immune system was not observed in subsequent litters from the same dams, supporting the idea that the ability of the dams to enhance offspring immunity was not caused by long-term maternal imprinting (Fig. 1E). Thus, a transient infection during pregnancy can impose long-lasting, tissue-specific alterations to offspring immunity.

We next assessed whether this phenomenon was T cell intrinsic or caused by an altered tissue environment. To this end, we evaluated the ability of the offspring delivered by control or previously infected dams to mount de novo responses to the microbiota. Transgenic T cells specific for a defined commensal-derived antigen (CBir1 Tg cells) (*22*) were transferred into the offspring of naïve or previously infected dams (Fig. 1F). Consistent with previous studies (*22*, *23*), a small number of CBir1 Tg cells accumulated in the siLP of control mice, where they remained largely undifferentiated (Fig. 1F). By contrast, a higher number of CBir1 Tg cells accumulated in the siLP of offspring delivered by previously infected dams, with a significant fraction of those differentiated toward a T_H17 cell phenotype (Fig. 1F). Thus, maternal infection can affect the offspring intestinal milieu in a way that enhances reactivity toward the microbiota and promotes T_H17 cell responses. In support of this observation, whole-tissue RNA sequencing of the small intestine revealed the up-regulation of genes encoding for factors known to promote local licensing of T_H17 cells (serum amyloid A, *Saa3*) (*24*) and of genes downstream of T_H17 cell responses (*S100a8*, and *S100a9*) (Fig. 1G).

Maternally restricted infection alters offspring intestinal immunity in utero independently of the microbiota

Infection can affect the composition, localization, and function of the microbiota (*25*–*27*). As such, we next explored the possibility that enhanced T_H17 cell responses could be associated with acquisition of infection-altered maternal microbiota. The microbiota of the offspring delivered by infected and naïve dams clustered closely to each other, as demonstrated by Bray–Curtis principal coordinates analysis (Fig. 2, A and B). Although no significant shifts occurred at the phylum level (Fig. 2C), a small number of taxa belonging to *Bacteroidales* and *Coriobacteriales* were increased, whereas a few taxa belonging to the *Clostridiales* and *Erysipelotrichales* were decreased, in the offspring of previously infected dams (Fig. 2D). To assess the potential contribution of the altered microbiota to offspring immunity, newborns delivered by control and previously infected dams were cross-fostered until weaning (Fig. 2E). Offspring delivered by previously infected dams and fostered by control dams harbored a number of siLP T_H17 cells comparable to those of offspring delivered and fostered by previously infected dams (Fig. 2, F to I). Conversely, offspring delivered by control dams and fostered by previously infected dams did not have an increased number of T_H17 cells within the siLP (Fig. 2, F to I). These data support the hypothesis that the heightened intestinal T_H17 cell responses caused by maternal infection were acquired in utero and did not result from the transfer of altered maternal microbiota after birth or during lactation.

Increasing maternal IL-6 levels during pregnancy is sufficient to enhance intestinal T_H17 cell responses and reactivity to the microbiota in offspring

Soluble factors, including cytokines, can cross the placental barrier (*28*, *29*). To assess a potential role for maternal soluble factors, we transferred sera from infected dams to naïve dams at gestation days 13.5 and 15.5 (Fig. 3A). We confirmed that no bacteria were transferred and that the recipient dams did not mount an immune response against *yopM* (fig. S2, A and B). Injection of sera from previously infected dams was sufficient to increase the offspring intestinal T_H17 cell number in a manner comparable to that observed after maternal *yopM* infection (Fig. 3A). Assessment of cytokine and chemokine levels in the sera of *yopM*-infected pregnant dams at different time points after infection revealed that four cytokines were significantly enriched: interleukin-6 (IL-6), interferon-γ (IFN-γ), granulocyte colony-stimulating factor (G-CSF), and tumor necrosis factor-α (TNF-α) (Fig. 3B). A single injection of IL-6, but not IFN-γ, G-CSF, or TNF-α, to naïve pregnant dams was sufficient to increase the number of T_H17 cells in the siLP of offspring (fig. S2C) in a dose-dependent manner (Fig. 3C). After injection, dams had circulating IL-6 levels that were comparable to those previously reported in the context of various infection models (*30*, *31*) and returned to baseline by day 3 after injection. Furthermore, IL-6 was not increased in the milk. In support of potential transplacental transfer, IL-6 levels were increased in fetal sera after both maternal *yopM* infection and IL-6 injection (fig. S2D).

IL-6 has been previously linked to various biological outcomes in the context of pregnancy, including gestational disorders (*32*). Furthermore, a previous study showed that IL-6 enhanced IL-17A production in dams, the production of which can contribute to the development of an autism-like disorder in offspring (*33*). However, in our present study, the enhanced IL-6 levels during pregnancy had no adverse impact on pregnancy and were not associated with detectable impairment in social preference of adult offspring (fig. S2E). One notable difference in experimental settings was that our study was conducted in mice devoid of segmented filamentous bacteria (SFB), a group of bacteria previously implicated in altered behaviors in progeny (*34*).

As observed in the context of maternal infection, IL-6 injection significantly increased the number of T_H17 cells in all sections of the offspring small intestine (Fig. 3D). IL-6 injection was also associated with a significant increase in the level of IL-17A production by T_H17 cells in the jejunum and the ileum (Fig. 3, E and F). Supporting enhanced tissue licensing of IL-17A signaling (*24*), both maternal infection and IL-6 injection enhanced serum amyloid A production in the jejunum and the ileum (Fig. 3G). IL-6 injection also induced a small increase in the number of $CD8^+$ T cells in the duodenum (fig. S3A). Droplet-based 3′ single-cell RNA sequencing (scRNA-seq) analysis revealed that siLP T_H17 cells of offspring from IL-6–injected dams expressed higher levels of canonical T_H17 cell markers (*Maf*, *Ahr*, *Ccr6*, *Il17a*, *Il17f*, and *Il22*) but not of genes associated with pathogenic T_H17 cells (*35*, *36*), including *Tbx21*, *Gmzb*, *Ifng*, and *Csf2* (Fig. 3, H and I). Thus, in agreement with the concept that maternal infection enhances T_H17 cell responses to the microbiota (Fig. 1F), T_H17 cells that accumulate under these conditions are nonpathogenic and likely involved in the control of barrier function. Indeed, we could not detect any differences in the distance between gut bacteria and host epithelium in offspring from IL-6–treated or infected dams (fig. S3, B and C). To address the possibility that IL-6 is not only sufficient but also necessary for the impact of maternal *yopM* infection on offspring immunity, pregnant dams were treated with a neutralizing IL-6 antibody before *yopM* infection (Fig. 3J). This treatment did not affect the ability of the pregnant dams to control the infection (fig. S3D). IL-6 neutralization during maternal infection significantly reduced the accumulation of T_H17 cells in the siLP of offspring (Fig. 3J). Collectively, our findings support the

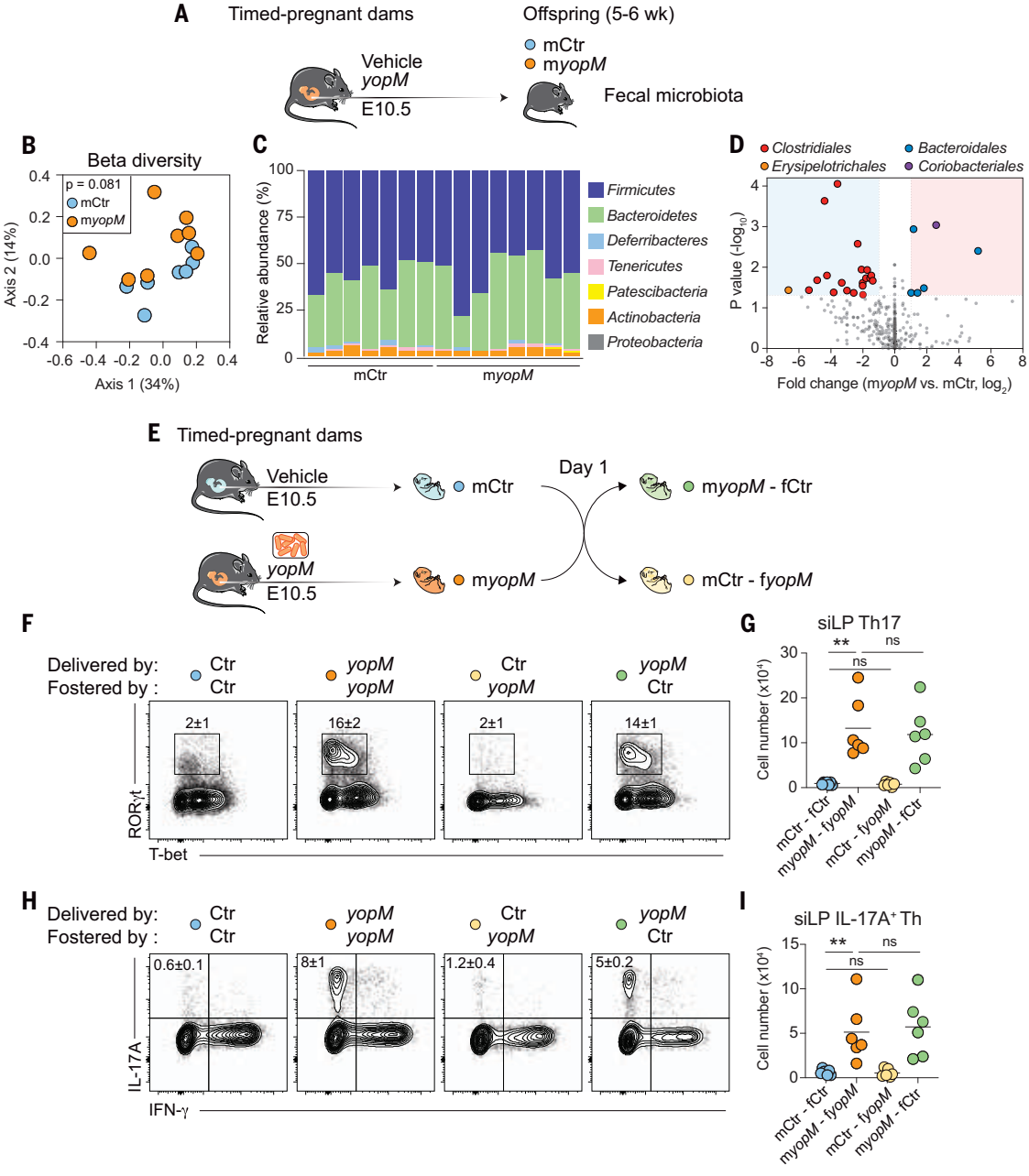


Fig. 2. Maternally restricted infection alters offspring intestinal immunity in utero independently of the microbiota. (**A**) Timed-pregnant dams were orally administered vehicle (PBS) or *yopM*, and fecal microbiota from their offspring, mCtr or m*yopM*, were profiled at 5 to 6 weeks of age by 16*S* rRNA gene sequencing. (**B**) Principal coordinates analysis plot displaying the Bray–Curtis distances between the fecal 16*S* profiles of mCtr and m*yopM* offspring. Percentages represent the variance corresponding to each principal coordinate. Significance was determined by PERMANOVA. (**C**) Relative abundance of phyla in the mCtr and m*yopM* fecal microbiota. (**D**) Volcano plot of fecal microbiota species (sequence variants). The taxonomic families of significantly altered species are indicated. (**E**) Cross-fostering experiment. Within 24 hours after birth, offspring delivered by vehicle-treated dams were transferred to previously infected dams (mCtr-f*yopM*) or vice versa (m*yopM*-fCtr). Offspring were cross-fostered until weaning and analyzed at 5 to 8 weeks of age. In the control groups, the offspring were raised by their own dams (mCtr-fCtr, m*yopM*-f*yopM*). (**F**) Representative contour plots of transcription factors expression by T_H cell subsets in the offspring siLP. (**G**) Total siLP T_H17 (RORγt$^+$) cell number. (**H**) Representative contour plots of cytokine production by T_H cells after restimulation. (**I**) Total cell number of siLP IL-17A–producing T_H cells. In (A) to (I), all flow plots were gated on live CD45$^+$CD90.2$^+$TCRβ$^+$CD4$^+$Foxp3$^-$. Numbers in representative flow plots indicate mean ± SD. Data are representative of three independent experiments, each experiment with one pregnant dam per group using three to six offspring per pregnancy. Each dot represents an individual mouse. $^{**}P < 0.01$; ns, not significant (one-way ANOVA multiple-comparisons test).

idea that increased levels of maternal IL-6, elicited by transient infection during pregnancy, alter offspring intestinal immunity.

Although a shift in the maternally acquired microbiota was not responsible for altered offspring immunity (Fig. 2, D and G), we assessed whether the maternal microbiota was required during pregnancy and/or after delivery for the remodeling of the offspring immune response. To address this point, germ-free pregnant dams were injected with IL-6 or vehicle control. IL-6 injection did not cause enhanced T_H17 cell accumulation in the offspring siLP if the offspring were maintained under germ-free conditions (Fig. 3K). However, reintroduction of the microbiota (conventionalization) after weaning was associated with enhanced T_H17 cell number in the offspring from IL-6–injected dams but not in those from control dams (Fig. 3K). Thus, prenatal establishment and postnatal maintenance of IL-6–mediated tissue imprinting is independent of the maternal microbiota but allows the offspring to mount enhanced T_H17 cell responses in response to postnatal exposure to the microbiota.

Increasing maternal IL-6 levels during pregnancy alters the epigenome of fetal IECs

Responses to IL-6 signaling are mediated by the IL-6 receptor (IL6RA, also known as CD126),

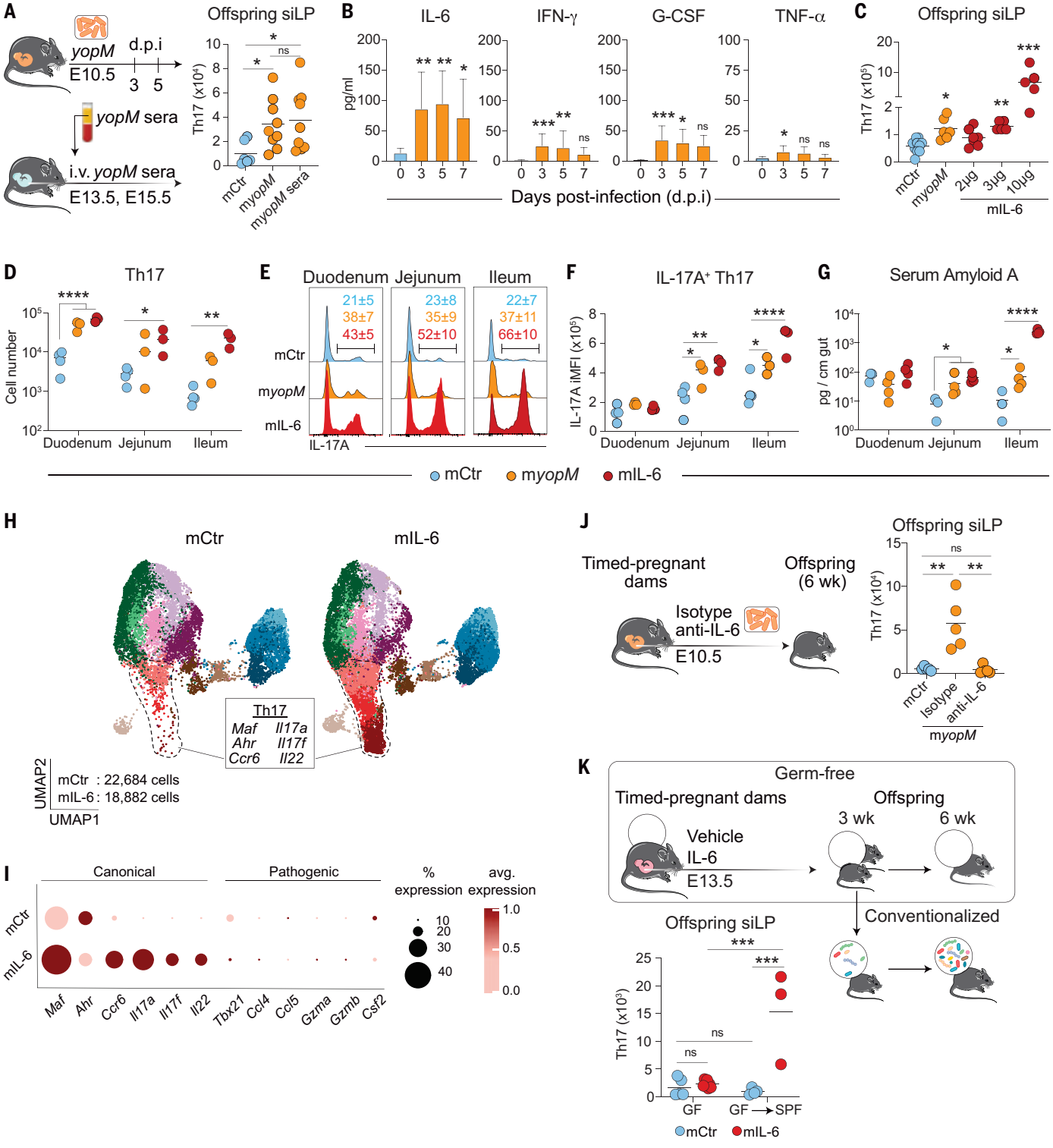


Fig. 3. Increasing maternal IL-6 levels during pregnancy is sufficient to enhance offspring intestinal T_H17 and responsiveness to the microbiota. (**A**) Left: Timed-pregnant dams were infected with *yopM*, and sera collected at 3 and 5 days postinfection (d.p.i.) were transferred intravenously to naïve pregnant dams at E13.5 and E15.5. Right: Total T_H17 cell number in the siLP of 6-week-old offspring. (**B**) Timed-pregnant dams were orally infected with *yopM* at E10.5. Sera collected before infection or at 3, 5, or 7 d.p.i. were assayed for cytokine and chemokine levels. (**C**) Timed-pregnant dams were orally administered vehicle or *yopM* or intravenously injected with the indicated doses of IL-6, and their offspring (mCtr, m*yopM*, or mIL-6) were analyzed at 6 weeks of age for total siLP T_H17 cell number. (**D** to **G**) siLP subcompartments from adult offspring delivered by mCtr, m*yopM*, or dams intravenously injected with 10 μg IL-6 (mIL-6) were analyzed for total T_H17 cell number (D), frequency of IL-17A–producing T_H17 cells (E), integrated mean fluorescence intensity (iMFI) of IL-17A production by T_H17 cells [iMFI = (IL-17A MFI) × (frequency of IL-17A–producing T_H17 cells)] (F), and serum amyloid A levels (G). (**H** and **I**) Total $CD4^+$ T cells were sorted from the siLP of mCtr or mIL-6 at 6 weeks of age and analyzed by scRNA-seq. (H) UMAP projection plots showing the expression

profiles of CD4$^+$ T cells. Colors represent cells clustered together on the basis of similarity of global gene expression. (I) Dot plot indicating level and abundance of expression of canonical or pathogenic gene signatures of the T_H17 cell cluster. (**J**) Left: *yopM*-infected dams were pretreated with 1 mg of anti-IL-6 monoclonal antibody or the isotype control 2 hours before infection. Right: Total T_H17 cell number in the siLP of 6-week-old offspring. (**K**) Top: Germ-free timed-pregnant dams were intravenously injected with vehicle or 10 μg of IL-6 at E13.5, and their offspring, mCtr or mIL-6, were maintained under germ-free conditions or conventionalized in a specific-pathogen-free facility from 3 to 6 weeks of age. Bottom: total T_H17 cell number in the siLP of 6-week-old offspring. In (A) to (K), data are representative of three independent experiments [(A) to (G) and (J) to (K)], each experiment with one or two pregnant dams per group using three to five offspring per pregnancy, or one experiment [(H) and (I)] with three pregnant dams per group using one offspring per pregnancy. Each dot represents an individual mouse. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; ns, not significant [ANOVA multiple-comparisons test, one-way in (A) to (C) and (J) and two-way in (D) to (G) and (K)]. See also figs. S2 and S3.

the glycoprotein 130 complex, and the JAK-STAT3 pathway (*37*). IL-6 has also been shown to cross both human and murine placenta (*28*, *38*, *39*). To investigate the impact of IL-6 on the fetal intestine, we first assessed expression of IL6RA in fetal intestinal epithelial and hematopoietic cells. We found that IL6RA was expressed at an intermediate level in all fetal intestinal epithelial cells (IECs) (EpCAM$^+$) and at a high level in a fraction of hematopoietic cells (CD45$^+$) (Fig. 4, A and B). By contrast, IL6RA was not expressed by fetal lung epithelial cells (EpCAM$^+$) or fetal skin keratinocytes (CD49f$^+$) (Fig. 4C). Of relevance to human settings, reanalysis of a published dataset (*40*) revealed that *IL6R* is also expressed at intermediate levels by human fetal IECs (fig. S4A). To identify cellular compartments that can respond to IL-6, we evaluated STAT3 phosphorylation at days 1 and 3 after *yopM* infection or IL-6 injection. A small fraction of IECs expressed phosphorylated STAT3 (pSTAT3) at steady state, supporting the idea that the JAK-STAT3 pathway may already be constitutively engaged during epithelial cell development (Fig. 4D and fig. S4B). At day 3 after maternal infection or IL-6 injection, pSTAT3 levels were further increased in fetal IECs but not in hematopoietic cells (Fig. 4D).

To test the possibility that IECs may be the direct target tissue imprinted by maternal IL-6 during infection, we characterized the epigenetic and transcriptional landscape of fetal IECs from control and IL-6–injected dams using an assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) and RNA-seq (fig. S4C). Genes important for IEC identity (including *Epcam*, *Vil1*, *Cdx1*, and *Cdx2*) (*41*, *42*) were equally accessible in fetal IECs from control and IL-6–injected dams (fig. S4D). Open chromatin regions in fetal IECs were enriched in transcription factor binding motifs (fig. S4E) that were previously implicated in intestinal development (CDX1/2), IEC differentiation (GATA and HNF4G), and epithelial cell identity establishment (KLF) (*43*). However, on a genome-wide level, fetal IEC chromatin accessibility was substantially increased after injection of IL-6 during pregnancy, with 2118 regions more accessible in fetuses from IL-6–injected dams (greater than a twofold change), compared with only four regions in IECs from fetuses from control dams (Fig. 4E). The differentially accessible promoter regions revealed enrichment of genes associated with intestinal physiological functions, including peptide transport, stem cell division, epithelium development, and T_H cell differentiation (e.g., *Jak2* and *Il10rb*) (Fig. 4F). By contrast, transcriptomic analysis showed minor alterations in fetal IECs (fig. S4F). Thus, exposure to IL-6 during development increases global chromatin accessibility of fetal IECs with minor impacts on their transcriptional landscape.

To further assess the contribution of IL-6 signaling on the ability of fetal IECs to regulate intestinal immunity, we specifically deleted the IL-6 receptor gene *Il6ra* from IECs (Fig. 4G). To this end, we used a breeding scheme allowing IL-6 signaling to be intact in the dams (*Il6ra*$^{fl/fl}$ *Vil1*$^{Cre-}$) and for IECs of the offspring to be either *Il6ra* sufficient (*Il6ra*$^{fl/fl}$ *Vil1*$^{Cre-}$) or *Il6ra* deficient (*Il6ra*$^{fl/fl}$ *Vil1*$^{Cre+}$). Confocal imaging confirmed selective ablation of IL6RA in IECs in *Il6ra*$^{fl/fl}$ *Vil1*$^{Cre+}$ mice (fig. S5, A and B). After injection of IL-6 during pregnancy, intestinal T_H17 cells were significantly reduced in the gut of *Il6ra*$^{fl/fl}$ *Vil1*$^{Cre+}$ offspring compared with *Il6ra*$^{fl/fl}$ *Vil1*$^{Cre-}$ littermate controls (Fig. 4G). Thus, IL-6 signaling in fetal IECs is sufficient and necessary to confer long-term intestinal immune alterations.

Prebirth IL-6 exposure alters chromatin accessibility and transcriptome of intestinal epithelial stem cells

Prior lineage-tracing studies revealed that all fetal IECs can contribute to the adult intestinal epithelial stem cell (EpSC) pool that is characterized by *Lgr5* expression (*44*). To address the possibility that alterations in fetal IECs may have long-term consequences on adult EpSCs, we assessed the chromatin accessibility of EpSCs (EpCAM$^+$Lgr5$^+$) from 6-week-old offspring delivered by control or IL-6–injected dams using ATAC-seq (Fig. 5A and fig. S6A). Chromatin accessibility was increased in EpSCs of adult offspring delivered by IL-6–injected dams, with 1006 regions more open (greater than a twofold change), compared with only 71 regions in EpSCs of control offspring (Fig. 5B). As anticipated based on the profound changes in epithelial cells occurring after birth and during weaning (*45*, *46*), open chromatin sites were largely distinct from those observed in the fetuses, with only 17 common peaks (<1% of the fetus) shared between fetal IECs and adult EpSCs (fig. S6B). Differentially accessible promoters found in adult EpSCs of offspring from IL-6–injected dams showed enrichment of genes involved in MAPK activation, cytokine signaling, and defense responses (e.g., *Alox5ap*, *Ptgs2*, *Mgst2*, and *Ffar3*) (Fig. 5C).

In adults, intestinal EpSCs constantly self-renew and differentiate into diverse subsets of epithelial cells endowed with discrete functions (*47*). Thus, we used droplet-based 3′ scRNA-seq to evaluate the potential consequences of maternal IL-6–mediated EpSC alterations on the composition and transcriptome of IECs (Fig. 5D). We focused our analysis on the duodenum and ileum, where we observed the largest increase in T_H17 cells after maternal IL-6 injection (Fig. 3D). Unsupervised clustering partitioned the IECs isolated from offspring delivered by control (n = 3, 4929 cells) and IL-6–injected (n = 3, 4112 cells) dams into 21 clusters (Fig. 5E and fig. S7A). Each cluster was identified (fig. S7B) on the basis of signature genes associated with distinct cell types and differentiation states (*48*). No significant differences were observed in cell composition between offspring delivered by control and IL-6–injected dams (Fig. 5F). By contrast, IL-6 injection imposed discrete transcriptomic alterations in defined subsets of IECs, including stem cells, transit-amplifying cells, and subsets of enterocytes at various stages of differentiation (fig. S7C). Genes encoding canonical components of the major histocompatibility complex II (MHCII) machinery, including *Cd74*, *H2-Aa*, and *H2-Ab1*, were strongly enriched in both stem cells and enterocytes of offspring delivered by IL-6–injected dams compared with control dams, with the strongest differences observed in the duodenum (Fig. 5G and fig. S7C). The expression of MHCII and genes encoding for the antigen presentation machinery by EpSCs has recently been reported to be microbiota dependent and to contribute to epithelial cell differentiation (*49*, *50*). IL-6 injection during pregnancy was also associated with increased expression of genes encoding antimicrobial peptides (e.g., *Reg3b* and *Reg3g*) in enterocytes (Fig. 5G). Thus, an increase in IL-6 during pregnancy can have a long-lasting impact on EpSC chromatin accessibility and downstream epithelial cell function.

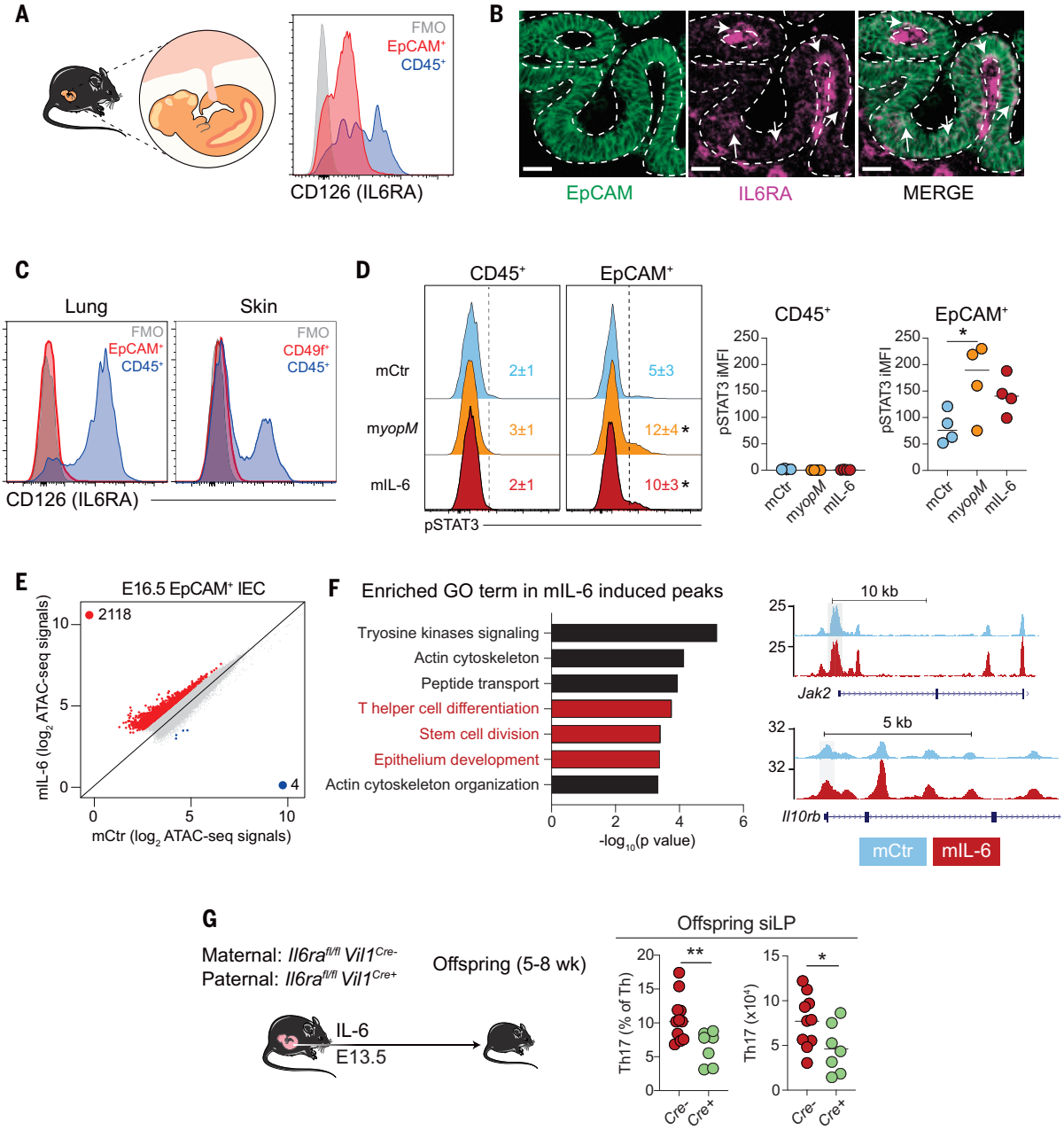


Fig. 4. Increasing maternal IL-6 levels during pregnancy alters the epigenome of fetal IECs. (**A**) Left: Fetal intestines were assessed for IL6RA (CD126) expression. Right: Representative histogram showing IL6RA expression by fetal IECs (IECs, $CD45^{-}EpCAM^{+}$) and hematopoietic cells ($CD45^{+}$) at E15.5. FMO, fluorescence minus one control. (**B**) Representative micrograph of fetal intestines stained for EpCAM and IL6RA at E15.5. Arrows point to IL6RA expressed by $EpCAM^{+}$ IECs (denoted by dashed line). Scale bars, 30 μm. (**C**) Representative histogram showing IL6RA expression by fetal lung epithelial cells ($CD45^{-}EpCAM^{+}$), fetal skin keratinocytes ($CD45^{-}CD49f^{+}$), and hematopoietic cells ($CD45^{+}$) at E15.5. (**D**) Left: Representative histogram of pSTAT3 expression by fetal hematopoietic cells or IECs from dams at 3 days after treatment with vehicle (mCtr), *yopM* (m*yopM*), or 10 μg IL-6 (mIL-6). Right: iMFI of pSTAT3 in hematopoietic cells ($CD45^{+}$) or IECs ($EpCAM^{+}$). (**E** and **F**) ATAC-seq data of fetal IECs from dams 3 days after treatment with vehicle (mCtr) or intravenous injection with 10 μg of IL-6 (mIL-6). (E) Scatter plot of all accessible chromatin regions. Red and blue dots correspond to regions that were differentially accessible by more than twofold in the IECs of mIL-6 fetuses and mCtr fetuses, respectively. (F) Left: Top Gene Ontology (GO) terms that were enriched in differentially accessible promoter regions (peaks within 3 kb of the next transcription start site, fold change >2) of mIL-6 fetuses compared with mCtr fetuses. Right: UCSC Genome Browser snapshots of genes of interest. Promoters open in mIL-6 offspring are highlighted in gray boxes. (**G**) *Il6ra* was deleted from IECs by crossing $Il6ra^{fl/fl}$ $Vil1^{Cre-}$ females to $Il6ra^{fl/fl}$ $Vil1^{Cre+}$ males. Dams were injected with 10 μg of IL-6, and offspring were analyzed at 5 to 8 weeks of age for siLP T_H17 cells. In (A) to (G), data are representative of three independent experiments [(A) to (D) and (G)], each experiment with one to three pregnant dams per group, or one experiment [(E) and (F)] with three pregnant dams per group using four to seven offspring per pregnancy. $^{*}P < 0.05$; $^{**}P < 0.01$ [two-tailed unpaired Student's *t* test in (G) and one-way ANOVA multiple-comparisons test in (D)]. See also figs. S4 and S5.

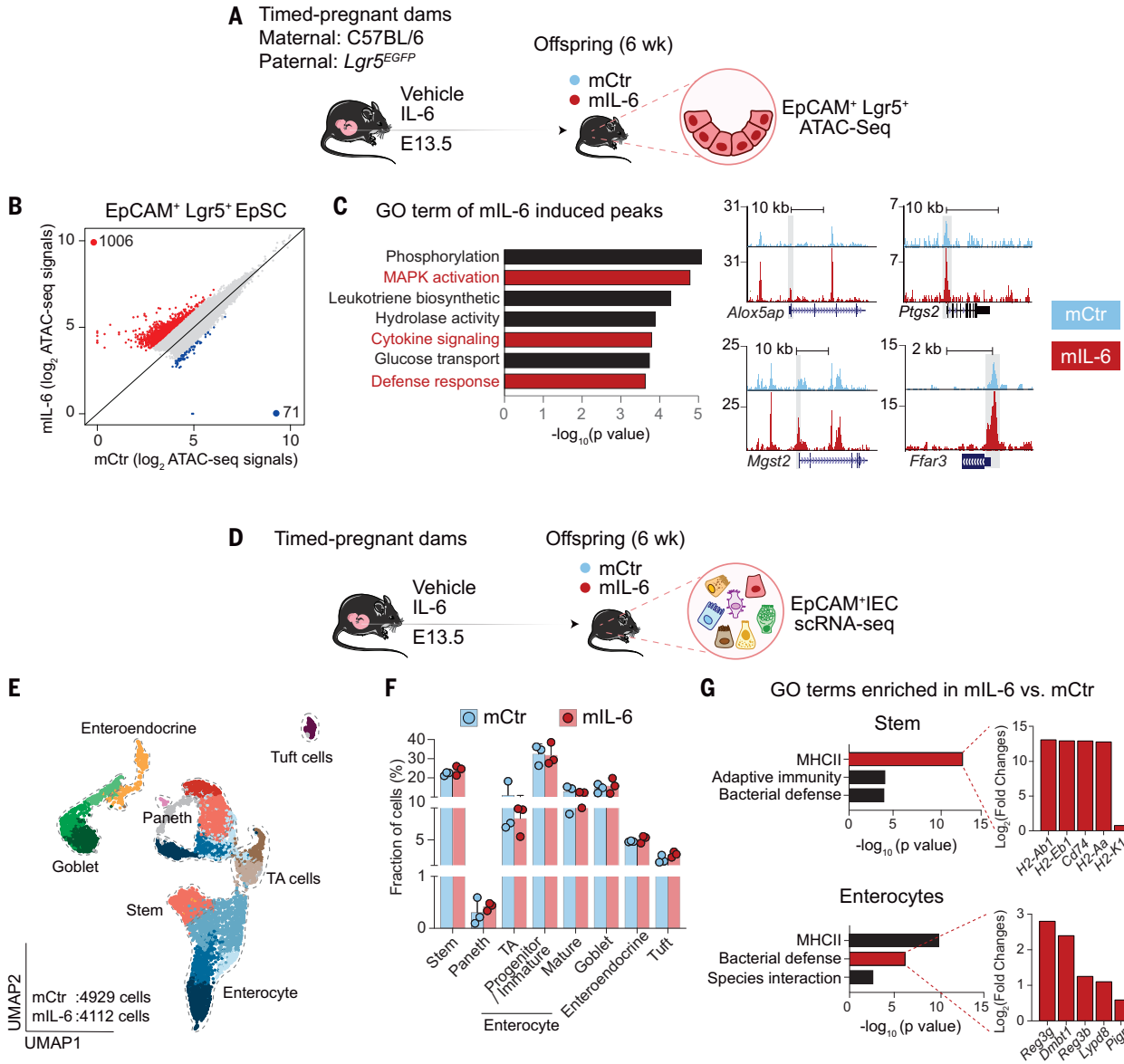


Fig. 5. Increasing maternal IL-6 levels during pregnancy alters the chromatin accessibility and transcriptome of offspring intestinal epithelial stem cells. (**A**) Timed-pregnant dams were injected with vehicle or 10 μg of IL-6 at E13.5. Intestinal EpSCs were sorted from their offspring, mCtr or mIL-6, for ATAC-seq at 6 weeks of age. (**B**) Scatter plot showing open chromatin regions. Red and blue dots correspond to regions that were differentially accessible by more than twofold in the EpSCs of mIL-6 and mCtr offspring, respectively. (**C**) Left: GO term analysis of differentially accessible promoter regions (peaks within 3 kb of the next transcription start site, fold change >2) in mIL-6 EpSC compared with mCtr EpSC. Right: UCSC Genome Browser snapshots of genes involved in cytokine signaling. Promoters open in mIL-6 offspring are highlighted in gray boxes. (**D**) IECs were sorted from mCtr or mIL-6 offspring for scRNA-seq at 6 weeks of age. (**E**) UMAP projection of IEC subsets, assigned on the basis of the 25 most differentially expressed genes in each subset versus all other cells. (**F**) Frequency of each IEC subset. (**G**) The top three GO terms that were enriched in stem cells and enterocytes of mIL-6 offspring. For selected GO terms highlighted in red, the top five genes up-regulated in mIL-6 offspring are shown. Data are from one experiment with three pregnant dams per group using one offspring per pregnancy. See also figs. S6 and S7.

Maternal infection or IL-6 exposure promotes offspring intestinal protective immunity and inflammation

Altered epithelial activation status and enhanced T cell responses to the microbiota in the gut of offspring of previously infected or IL-6–injected dams pointed to the possibility that this phenomenon may be associated with enhanced antimicrobial defense. To address this hypothesis, we used an acute model of oral infection with *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium) (Fig. 6A). At the dose and route of infection used, *S.* Typhimurium primarily invades IECs before disseminating to distal organs and causing host lethality (*51*). Although no differences in bacterial shedding were observed at days 1 and 2 after infection, by day 4, offspring of IL-6–injected dams displayed a 1- to 2-log reduction in bacterial burden in their feces and Peyer's patches compared with the offspring of control dams (Fig. 6, B and C). Enhanced IL-6 during pregnancy also reduced bacterial dissemination to distal organs in offspring, with a 2- to 3-log decrease in bacterial burden in the mesenteric lymph nodes, liver, and spleen compared with controls (Fig. 6C) and significantly improved offspring survival rates after infection (Fig. 6D). Consistent with enhanced control of the infection, lipocalin-2, a marker

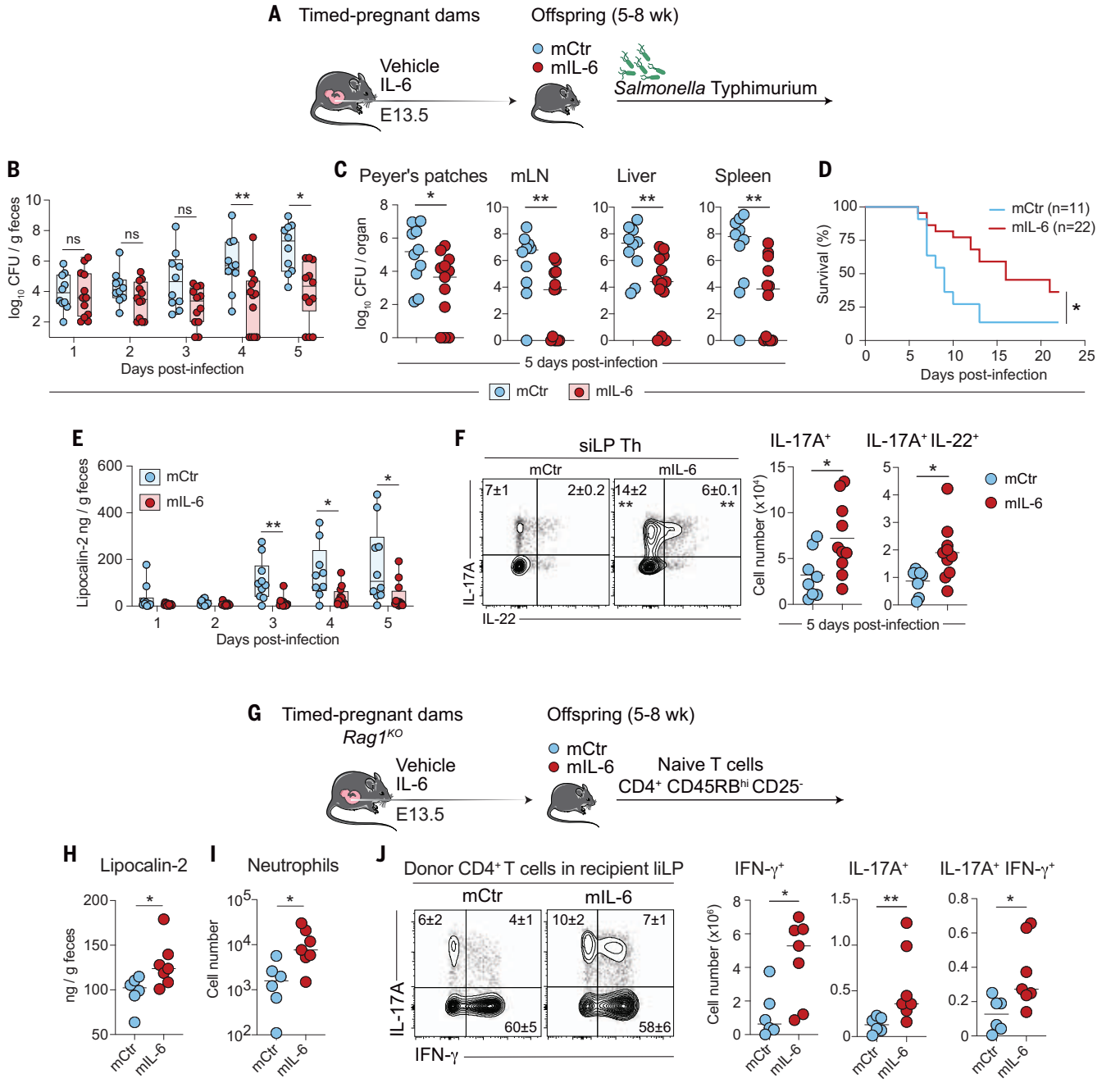


Fig. 6. Maternal infection or IL-6 exposure promotes offspring intestinal protective immunity and inflammation. (**A**) Timed-pregnant dams were injected with vehicle or 10 μg of IL-6 at E13.5, and their offspring (mCtr or mIL-6) were orally administered *S.* Typhimurium at 5 to 8 weeks of age. (**B** and **C**) *S.* Typhimurium burden in feces at days 1 to 5 after infection (B) and in tissues at day 5 after infection (C). (**D**) Survival curve. (**E**) Fecal lipocalin-2 at days 1 to 5 after infection. (**F**) Left: Representative contour plots showing cytokine production by siLP T_H cells ($CD45^+CD90.2^+TCR\beta^+CD4^+Foxp3^-$). Right: Total IL-17A$^+$ and IL-17$^+$IL-22$^+$ siLP T_H cells. (**G**) Timed-pregnant *Rag1*^{KO} dams were intravenously injected with PBS or 10 μg of IL-6, and their offspring at 5 to 6 weeks of age (mCtr or mIL-6) were intraperitoneally injected with sorted 4×10^5 naïve T cells ($CD4^+CD45RB^{hi}CD25^-$) from wild-type donors. (**H** and **I**) Intestinal inflammation was assessed at 6 to 8 weeks after transfer by fecal lipocalin-2 (H) and total neutrophil numbers (I) in liLP. (**J**) Left: Representative contour plots of cytokine production by donor CD4$^+$ T cells in liLP after restimulation. Right: Total CD4$^+$ T cell numbers of IFN-γ, IL-17A, and double producers present in the liLP. In (B) and (C), the CFU detection limit was 10^2 CFU per gram of organ. Numbers in representative flow plots indicate mean ± SD. Data are representative of two or three independent experiments with one or two pregnant dams per group using three to six offspring per pregnancy. Each dot represents an individual mouse. *P < 0.05; **P < 0.01; ns, not significant [two-tailed unpaired Student's t test in (B), (C), (E), (F), and (H) to (J) and a log-rank test in (D)]. See also figs. S8 and S9.

of mucosal inflammation (52), was also significantly reduced in the offspring of IL-6–injected dams compared with controls (Fig. 6E). Further, antimicrobial peptides, such as *Reg3b* and *Pigr*, were significantly increased in IEC of offspring delivered by IL-6–injected dams at 2 days after infection (fig. S8A). Although the direct role of IL-17A in controlling *S.* Typhimurium remains unclear (53), infection with this bacterium has been associated with early type 3 immune responses (54). At day 5 after infection, the number and frequency of IL-17A- and IL-17A/IL-22–producing $CD4^+$ T cells, but not of ILCs, were significantly increased in the siLP of offspring of IL-6–injected dams compared with control dams (Fig. 6F and fig. S8B). Protection afforded by IL-6 injection in dams was significantly reduced in mice for which *Il6ra* was specifically deleted from IECs (fig. S8, C to E). Furthermore, offspring delivered by *yopM*-infected dams also displayed enhanced control of *S.* Typhimurium infection in a manner comparable to what was observed after IL-6 injection (fig. S8, F to H). Thus, both maternal infection and enhanced IL-6 signaling experienced by the fetus during pregnancy can promote antimicrobial defense through its direct impact on epithelial cells. No differences were observed in the ability of offspring delivered by IL-6–injected dams to control cutaneous infection with *Candida albicans*, which further supports the concept that, in this context, maternal imprinting of offspring is restricted to the gut compartment (fig. S9, A to C).

Heightened immunity can also come at the cost of increased susceptibility to inflammatory disorders. To test the possibility that maternal infection and/or IL-6 could contextually promote inflammatory responses in the offspring, we used two well-defined models of experimental colitis (55, 56). Using a naïve T cell transfer model, we found that offspring of IL-6–injected dams displayed enhanced intestinal inflammation compared with control offspring, as evidenced by increased fecal lipocalin and a higher number of neutrophils in the liLP (Fig. 6, G to I). Additionally, the number of transferred T cells producing IFN-γ, IL-17A, or both (IFN-γ^+ IL-$17A^+$) were significantly increased in the liLP of offspring from IL-6–injected dams compared with control offspring (Fig. 6J). The significant increase in double producers is of particular interest because this specific coexpression of cytokines is one of the characteristics of pathogenic T_H17 cells (57). Additionally, offspring of *yopM*-infected or IL-6–injected dams displayed enhanced pathology to dextran sulfate sodium (DSS)–induced colitis compared with controls, as evidenced by increased bodyweight loss and colon shortening (fig. S9, D to F). Thus, heightened intestinal immunity caused by maternal infection may also come at a cost of enhanced predisposition to inflammatory disorders.

Discussion

Optimal responses to microbial challenges are of utmost importance for host survival. Here, we showed that pregnancy represents a critical stage for tissue-specific immune education of the offspring. We demonstrated that direct responses of fetal epithelial cells to cytokines produced during maternal infections confer an enduring epigenetic memory to gut stem cells, a phenomenon that we propose contributes to both enhanced immunity and predisposition to inflammatory disorders. Recent studies have uncovered that, in adults, epithelial stem cells can develop memory of inflammatory insults, a phenomenon associated with accelerated responses to subsequent injuries (2, 3, 58). Our present work proposes that even a transient, mild infection encountered during prenatal development can impose lasting alterations to gut epithelial stem cells before birth, resulting in an altered tissue threshold of activation.

The impact of maternal infection was tissue specific and dominantly mediated by a single cytokine, IL-6. This observation raises an intriguing possibility that during fetal development, stem cells residing in different compartments may be educated by highly specific signals to develop optimal tissue immunity. Our findings also support the idea that infections experienced by the mother during pregnancy may optimize responses to immediate microbial threats in the offspring by augmenting responses at the target infectious sites. Further, IL-6 has previously been reported to be increased throughout pregnancy (59), a phenomenon that may broadly contribute to the education of the mammalian gut immune system.

Although our data indicate that maternal microbial exposure can be coopted by the fetus to develop optimal immune fitness, this phenomenon may also help to explain the marked increase in inflammatory disorders observed in the human population. Indeed, the altered maternal environment, together with the removal of regulatory microbial partners such as worms, may have contributed to shifting the outcome of maternal-induced immune maturation from protective to inflammatory. Further exploration of the maternal-fetal dialog in the face of microbial or environmental challenges may provide additional insights into our understanding of homeostatic and protective tissue immunity, as well as host predispositions to inflammatory diseases.

Materials and Methods

Mice

Specific pathogen–free (SPF) C57BL/6, *Lgr5*$^{EGFP\text{-}IRES\text{-}creERT2}$ (B6.129P2-*Lgr5*$^{tm1(cre/ERT2)/Cle}$/J, Jax008875), *Il6ra*$^{fl/fl}$ (B6.SJL-*Il6ra*$^{tm1.1Drew}$/J, Jax012944), and *Vil1*^{cre} (B6.Cg-Tg(Vil1-cre)1000Gum/J, Jax021504) mice were purchased from The Jackson Laboratory and Taconic Biosciences (SFB-negative barrier). Germ-free C57BL/6 mice were bred and maintained in the National Institute of Allergy and Infectious Diseases (NIAID) Microbiome Program gnotobiotic animal facility. CBir1 Tg mice were generated by Dr. C. Elson (University of Alabama, Birmingham), obtained under a material transfer agreement, and back-crossed to CD45.1-expressing mice (22). All mice were bred and maintained at an American Association for the Accreditation of Laboratory Animal Care (AAALAC)–accredited animal facility at NIAID and housed in accordance with procedures outlined in the *Guide for the Care and Use of Laboratory Animals*. All experiments were performed at NIAID under an animal study proposal (LISB19E and LHIM2E) approved by the NIAID Animal Care and Use Committee. For timed-pregnant breeding, males were first singly housed for 3 days and then cohoused with females for breeding. Males were separated from the breeding cage once a plug was observed. Littermate dams were used in each experiment to help normalize for differences in the microbiota. Offspring were weaned at 3 weeks after birth. Age-matched mice 5 to 8 weeks old from both sexes were used in each experiment unless specified otherwise.

Yersinia pseudotuberculosis infection

The *yopM* mutant strain of *Y. pseudotuberculosis* 32777 was grown overnight in 2× YT medium (Sigma-Aldrich) at 25°C with shaking at 200 rpm. Overnight cultures were centrifuged at 3000*g*, resuspended in phosphate-buffered saline (PBS), and adjusted to a density [absorbance at 600 nm (A_{600})] of 0.1 in PBS. Mice fasted overnight (12 to 16 hours) in new cages were orally administered 200 μl of *yopM* suspension [~2×10^7 colony-forming units (CFUs)]. To determine bacterial burden in the maternal spleen, liver, placenta, and fetal liver after *yopM* infection, tissues were homogenized through 70-μm filters with PBS. Tissue homogenates were serially diluted and plated onto MacConkey plates, and colonies were counted after incubation at 25°C for 48 hours. To measure bacterial burden in feces and fetal intestines, fecal DNA was extracted using phenol-chloroform, and intestinal DNA was extracted using TRIzol Reagent (Invitrogen). The *yscF* gene was quantified against a standard curve by quantitative PCR (forward primer 5′-ATGAGTAACTTCTCTG-GATTTACG-3′, reverse primer 5′-TTATGG-GAACTTCTGTAGGATG-3′).

Salmonella Typhimurium infection

A nalidixic acid-resistant strain of *S.* Typhimurium, IR715 (provided by Prof. A. Bäumler, University of California, Davis) was cultured in LB at 37°C with shaking at 200 rpm. Overnight cultures were centrifuged at 3000*g*, resuspended in PBS, and adjusted to an A_{600} of 0.001. Mice were fasted overnight (12 to 16 hours) before oral administration of 200 microliters of *S.* Typhimurium

suspension (~2×10^5 CFUs). Fecal samples were collected into preweighed tubes, homogenized in 500 μl of PBS with pipette tips, and vortexed for 5 s. At day 5 after infection, collected tissues were homogenized through 70-μm filters with PBS. The top fraction of fecal and tissues homogenates was serially diluted in PBS, spotted with a multichannel pipette onto LB agar with 50 mg/ml nalidixic acid, and incubated overnight at 37°C before counting CFUs. Remaining fecal homogenates were centrifuged at 3000*g* and stored at −80°C for lipocalin-2 measurement.

Adoptive transfer of CBir1 Tg CD4⁺ T cells

Cells from the spleen and the lymph nodes of CD45.1 CBir1 Tg mice were isolated by being passed through 70-μm filters. To enrich for $CD4^+$ T cells, single-cell suspensions were incubated with biotin-conjugated anti-CD8, anti-NK1.1, anti-CD11b, and anti-CD19 antibodies followed by anti-biotin microbeads (Miltenyi Biotec) and negatively selected by magnetic cell separation with MACS technology (Miltenyi Biotec). Then, 0.5 to 1.0×10^6 enriched $CD4^+$ T cells from CBir1 Tg mice were intravenously transferred by retroorbital injection to offspring delivered by control or transiently infected dams on a congenic CD45.2 background. Recipient mice were sacrificed at day 10 to 12 after transfer to assess donor cells in the siLP.

Cross-fostering experiments

Within 24 hours after birth, whole litters were removed from the original mother and gently mixed with new bedding in clean cages. Foster mothers were held to urinate on the pups. The cages were covered to avoid disturbances and monitored for at least 48 hours. Pups were nurtured by foster mothers until weaning at 3 weeks after birth.

Serum transfer experiments

Whole blood was collected from *yopM*-infected timed-pregnant dams at days 3 and 5 after infection. Blood samples were incubated at room temperature for 30 min and centrifuged at 2000*g* for 10 min. Sera were collected and transferred immediately by retroorbital injection to naïve timed-pregnant dams at gestation days 13.5 and 15.5. Each recipient received 200 μl of sera, combined from at least two mice, at each time point. Fifty microliters of serum collected from each mouse was plated onto MacConkey plates to assess CFUs.

In vivo cytokine stimulation and blockade

Timed-pregnant mice were intravenously injected with up to 10 μg of recombinant mouse cytokines (BioLegend) at gestation day 13.5. For IL-6 blockade experiments, timed-pregnant dams at gestation day 10.5 were injected intraperitoneally with either 1 mg of anti–IL-6 (MP5-20F3, BioXCell) or the IgG1 isotype control (HPRN, BioXCell) 1 hour before *yopM* infection.

Germ-free mice conventionalization

Offspring delivered by PBS or IL-6–injected germ-free timed-pregnant mice were kept in germ-free isolators and fostered by the same mothers until weaning. Subsequently, half of the litter was kept in isolators, whereas the other half was transferred using aseptic technique to a SPF facility at 3 weeks of age and analyzed at 5 to 8 weeks of age.

DSS-induced colitis

Offspring delivered by PBS-injected, *yopM*-infected, or IL-6–injected dams were provided drinking water with 2.5% DSS (ThermoFisher Scientific) for 7 days and allowed to recover for an additional 7 days. Fresh DSS was provided every 2 days. Body weights were monitored daily, and mice were sacrificed at day 14 after treatment.

Naïve CD4⁺ T cell transfer–induced colitis

$Rag1^{KO}$ females were bred with $Rag1^{KO}$ males. At day E13.5, timed-pregnant dams were injected with PBS or 10 μg of IL-6. Offspring delivered by PBS or IL-6–injected dams were intraperitoneally injected with 4×10^5 $CD4^+CD25^-CD45RB^{hi}$ T cells sorted from naïve mice. Mice were sacrificed at weeks 6 to 8 after transfer.

Cutaneous C. albicans infection

From a freezer stock, *C. albicans* (strain SC5314) was streaked onto a yeast extract peptone dextrose (YPD) agar plate and incubated at 37°C for 2 days. A single colony was cultured in YPD broth containing penicillin-streptomycin for 16 to 24 hours at 30°C. After two serial passages, *C. albicans* was centrifuged at 400*g* for 5 min, washed twice with sterile PBS, and reconstituted in PBS at 2×10^{10} cells/ml for infection. Mice were first anesthetized with a mixture of ketamine and xylazine (100 and 10 mg/kg of body weight, respectively). The stratum corneum was removed using a Dremel rotary tool for 10 s, and 10 μl of PBS or *C. albicans* was applied to the skin. Mice were sacrificed on days 2 and 5 after infection. Whole ears were removed, weighed, and homogenized in sterile medium before plating on BBL CHROMagar Candida Medium (BD Biosciences).

Milk collection

Dams at day 10 after parturition were separated from pups for 2 hours and anesthetized with a mixture of ketamine and xylazine (100 and 10 mg/kg of body weight). Dams were then intraperitoneally injected with 2 IU of oxytocin (Sigma-Aldrich). Milk was collected from mammary glands using a P10 pipette and stored at −80°C for subsequent analysis.

Tissue processing

Mice were euthanized with CO_2, perfused through the left ventricle of the heart with 10 ml of PBS, and the siLP, liLP, lung, and skin were collected and placed into cold complete medium (RPMI 1640 supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 1 mM nonessential amino acids, 20 mM HEPES, 50 mM β-mercaptoethanol, 100 U/ml penicillin, and 100 mg/ml streptomycin). For siLP and liLP preparation, the Peyer's patches and the mesenteric adipose tissue were removed. Tissues were opened, washed in cold PBS to remove feces, cut into 1- to 2-cm segments, and treated with complete medium containing 5 mM EDTA and 0.145 mg/ml dithiothreitol for 20 min at 37°C with constant stirring. Tissues were shaken vigorously for 1 min and further digested with 10 ml of medium containing 500 mg/ml DNase I (Sigma-Aldrich) and 100 mg/ml Liberase TL (Roche) with continuous stirring at 37°C. Lungs were diced and incubated in 2 ml of prewarmed medium containing 500 mg/ml DNase I and 100 mg/ml Liberase TL for 25 min at 37°C. To isolate cells from the skin, ear pinnae were excised and separated into ventral and dorsal sheets, digested by placing the dermal side down in medium containing 500 mg/ml DNase I with 250 mg/ml Liberase TL, and incubated for 90 min at 37°C. Digested intestines, lung, and skin were passed through 70-μm cell strainers. Leukocytes were enriched by resuspension in 4 ml of 37.5% Percoll and centrifuged at 400*g* for 5 min. Cells were then washed with PBS before downstream analysis.

In vitro restimulation

To assess cytokine production potential, single-cell suspensions were restimulated in complete medium containing 10% fetal bovine serum, 50 ng/ml phorbol myristate acetate (Sigma-Aldrich), 5 mg/ml ionomycin (Sigma-Aldrich), and a 1:1000 dilution of GolgiPlug (BD Biosciences) for 2.5 hours at 37°C.

Flow cytometry analysis

The fluorophore-conjugated antibodies used are listed in table S1. For intracellular cytokine and transcription factor staining, after surface staining, cells were fixed and permeabilized using the Foxp3/Transcription Factor Staining Buffer Set (ThermoFisher Scientific) for at least 1 hour at 4°C and stained with fluorophore-conjugated antibodies for at least 1 hour at 4°C. All staining was performed in the presence of purified anti-mouse CD16/32 and purified rat gamma globulin. Dead cells were excluded using 4′,6-diamidino-2-phenylindol (DAPI, Sigma-Aldrich) or LIVE/DEAD Fixable Blue Dead Cell Stain Kit (Invitrogen Life Technologies). For phospho-STAT3 staining, after surface staining, cells were fixed with 4% paraformaldehyde at room temperature for 10 min and permeabilized

in prechilled methanol at −80°C for 1 hour. Cells were subsequently washed with PBS containing 0.5% Triton X-100 and 0.1% bovine serum albumin (BSA) and stained with a phycoerythrin-conjugated phospho-STAT3 antibody (Cell Signaling Technology) or the isotype control in the same buffer at 4°C. Cells were washed and analyzed immediately. Integrated mean fluorescence intensity reflects the total functional responses of a population determined by level of expression and frequency of positive cells (*60*).

16S rRNA gene profiling

Fecal DNA was extracted using the MagAttract PowerMicrobiome DNA/RNA kit (Qiagen). *16S* sequencing libraries were generated using 100 ng of purified DNA as template and the V4 region-targeting primers 515F (5′-GTGCCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACHVGGGTWTCTAAT-3′). Amplicons were quantified, pooled at equimolar concentrations, and sequenced on a MiSeq (Illumina) using the v3 MiSeq Reagent Kit (Illumina). QIIME 2 was used to analyze *16S* data as previously described (*27*, *61*).

Confocal microscopy

Fetal intestines were isolated using a dissecting microscope. Adult small intestines were flushed with cold PBS and prepared using the Swiss-roll technique. Samples were incubated overnight in fixation and permeabilization buffer (BD Biosciences), followed by dehydration in 30% sucrose before embedding in OCT compound (Sakura Finetek). Then, 20-μm sections were cut on a CM3050S cryostat (Leica) and adhered to Superfrost Plus slides (VWR). Frozen sections were treated with prechilled acetone for 30 min at −80°C and subsequently permeabilized and blocked in PBS containing 0.3% Triton X-100 and 1% BSA (Sigma-Aldrich). Samples were stained with anti-IL6RA (BD Biosciences) and anti-EpCAM (BioLegend) at 37°C for 30 min. For fluorescence in situ hybridization, colons were fixed in Methacarn solution at 4°C overnight, washed in 70% ethanol, and embedded in paraffin. Longitudinal sections (5 to 25 μm) were hybridized to a bacterial *16S* rRNA gene probe: [AminoC6+Alexa488] – GCTGCCTCCCGTAGGACT – [AmC7-Q+Alexa488] (Eurofins MWG Operon) and counterstained with DAPI. Images were captured on a Leica TCS SP8 confocal microscope equipped with HyD and PMT detectors and a 40× oil objective (HC PL APO 40×/1.3 oil). Images were analyzed using Imaris (Bitplane).

IEC isolation and sorting

Freshly isolated and longitudinally opened samples were washed twice with gut wash buffer [Hanks' balanced salt solution supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 1 mM nonessential amino acids, 20 mM HEPES, 100 U/ml penicillin, 100 mg/ml streptomycin, and 2% fetal calf serum (FCS)]. Tissues were cut into 1- to 2-cm segments and incubated with gut wash buffer containing 2 mM dithiothreitol in a 37°C water bath for 10 min. Tissues were vigorously shaken and subsequently incubated with fresh gut wash buffer containing 5 mM EDTA in a 37°C water bath for 7 min. EDTA-containing fractions were collected, and tissues were incubated with fresh EDTA for another 7 min. The two EDTA-containing fractions (enriched with crypts) were passed through 100-μm filters and centrifuged at 300*g* for 5 min. Cell pellets were digested with Dulbecco's modified Eagle's medium (DMEM) supplemented with 0.1 U/ml dispase (Life Technology) and 100 mg/ml DNase I for 8 min in a 37°C water bath. Digestion was neutralized with 2.5% FCS-DMEM, and the pellets were washed twice with gut wash buffer. Cell pellets were stained with antibodies against EpCAM, lineage (TCRβ, TCRγδ, CD8, and CD11b), CD31, and DAPI on ice for 30 min. For scRNA-seq, TotalSeq-A hashtag oligonucleotides (BioLegend) were included along with surface antibodies. IECs were sorted as the DAPI$^-$CD31$^-$Lineage$^-$EpCAM$^+$ population using a MA900 cell sorter (Sony) with a 100-μm chip.

Whole-tissue RNA-seq

Three centimeters of ileum was collected and extensively washed with PBS to remove feces. Samples were homogenized by bead-beating in tubes containing 2.8-mm beads and TRIzol reagent (ThermoFisher Scientific). RNA was extracted according to the TRIzol protocol. cDNA was synthesized with the Ultralow V4 kit (Clonetech), and sequencing libraries were subsequently prepared with the Nextera XT DNA prep kit (Illumina). RNA-seq reads were mapped with STAR to the mouse reference genome (mm10) with default parameters. Gene expression was measured with HOMER using analyzeRepeats.pl with parameters -condenseGenes and -tpm (for scatter plots) or -noadj (for differential gene expression analysis). Differential gene expression was assessed with HOMER's getDiffExpression.pl using DESeq2. Genes with a greater than a twofold change and a false discovery rate (FDR) smaller than 0.05 were considered differentially expressed.

FAST-ATAC-seq

Fast-ATAC-seq was performed according to a published protocol (*62*). Briefly, 10,000 fetal IECs (EpCAM$^+$) or adult EpSCs (EpCAM$^+$Lgr5$^+$) were sorted and pelleted by centrifugation at 500*g* for 5 min at 4°C. Supernatant was removed, and cell pellets were resuspended in 45 μl of transposase mixture (25 μl of Tagment DNA buffer, 2.5 μl of Tagment DNA enzyme, 0.5 μl of 1% digitonin, and 17 μl of H_2O) and incubated for 30 min at 37°C with agitation at 300 rpm. Tagmented DNA was purified using the QIAGEN MinElute Reaction Cleanup Kit, and purified DNA was eluted in 10 μl of elution buffer. Tagmented fragments were amplified with 10 to 15 PCR cycles on the basis of an amplification curve. After purification using the QIAGEN PCR cleanup kit, samples were sequenced single end on a NextSeq 550 (Illumina). ATAC-seq reads from three biological replicates were demultiplexed using bcl2fastq and mapped to the mouse reference genome (mm10) with STAR and default parameters. All biological replicates were run through the irreproducible discovery rate (IDR) framework, and only replicates that passed all filters (peak consistency between true replicates, peak consistency between pooled pseudoreplicates, and self-consistency for each replicate) and had an IDR threshold lower than 0.05 were kept. For final analysis, all peaks that passed IDR filter and were found in all replicates were used. In total, there were 24,938 peaks in adult EpSC cells from mCtr samples and 30,614 peaks in adult EpSC cells from mIL-6 samples. For analysis of fetal IECs, 49,787 peaks were found in mCtr samples and 43,813 in mIL-6 samples. For the IDR pipeline, open regions of chromatin were called using HOMER's findPeaks (with parameters -style factor, -center, -F 0, -L 0, -C 0, -fdr 0.01, -minDist 200, and -size 200) (*63*).

scRNA-seq and transcriptome analysis

EpCAM$^+$ IECs or CD4$^+$ T cells (2.5×10^4) were isolated from each sample using a Sony MA900 cell sorter. All samples were labeled with TotalSeqA hashtags antibodies (BioLegend), pooled, and encapsulated into droplets using a Chromium Single Cell Controller (10X Genomics), and libraries were prepared using Chromium Single Cell 3 Reagent Kits v3 (10X Genomics). The Controller was loaded with 2.5×10^4 cells per lane. mRNA was prepared following the 10X Genomics user guide, and the HTO library was prepared according to published guidelines (*64*). Libraries were sequenced on a NextSeq 550, with 10% of the lane occupied by the HTO library and 90% by the mRNA library. HTO was assigned to each single cell. Data were normalized, and lanes were combined using the scTransform function in Seurat. Data are displayed as uniform manifold approximation and projection (UMAP) using 25 dimensions. Intestinal cell subsets from each cluster were defined by the top 25 differentially expressed genes and identified on the basis of published datasets (*48*). Gene expression within each cluster was compared between offspring delivered by control or IL-6–exposed dams using Seurat.

RNA extraction, cDNA synthesis, and qPCR of sorted IECs

A total of 2×10^5 sorted IECs was lysed in RLT buffer, and total RNA was isolated using the

RNeasy Micro Kit (QIAGEN) according to the manufacturer's instructions. Total RNA was reverse-transcribed into cDNA using Omiscript RT Kit (QIAGEN) according to the manufacturer's instructions. qPCR was performed using TaqMan Gene Expression Assays on a Quantstudio 6 Flex Real-Time PCR System (Life Technologies). The housekeeping gene *Gapdh* was used to normalize critical threshold (CT) values. TaqMan primers are commercially available: *Gapdh* (Mm99999915_g1), *Reg3b* (Mm00440616_g1), *Reg3g* (Mm0044127_m1), *Pigr* (Mm01288210_g1), and *Dmbt1* (Mm00455996_m1). Analysis was conducted using the ΔCT method.

Three-chamber social behavior tests

Eight-week-old offspring were tested for sociability using three-chamber social tests. Mice were habituated in the center chamber for 10 min, followed by a 10-min habituation session with access to all three empty chambers. After the two habituation sessions, the 10-min sociability test was conducted. An inverted wire pencil cup (novel object) was placed in one of the side chambers and the other side chamber had an inverted wire pencil cup containing an age- and sexed-matched C57BL6 mouse (novel mouse). Frequency to enter and time in each chamber was analyzed using Noldus Ethovision software (Noldus Information Tech Inc. Leesburg, VA, USA).

Measurement of cytokines and chemokines

Sera cytokine and chemokine levels were quantified using the Cytokine & Chemokine Convenience 36-Plex Mouse ProcartaPlex Panel 1A (Invitrogen) according to the manufacturer's instructions.

Measurement of serum amyloid A

Three-centimeter intestinal segments from the duodenum, jejunum, and ileum were collected and extensively washed with PBS to remove feces. Samples were homogenized by bead-beating with 400 μl of RIPA lysis buffer (Chem Cruz) containing proteinase inhibitor in tubes containing 2.8-mm beads. Extracted samples were centrifuged at 10,000*g* for 10 min at 4°C to remove debris. Serum amyloid A levels were quantified using the Mouse Serum Amyloid A Quantikine ELISA Kit (R&D) according to the manufacturer's instructions.

Measurement of lipocalin-2

Lipocalin-2 levels in fecal samples were measured using the Mouse Lipocalin-2/NGAL Quantikine ELISA Kit (R&D Systems) according to the manufacturer's instructions.

Measurement of IL-6

IL-6 levels in serum and breast milk were measured using the Mouse IL-6 Quantikine ELISA Kit (R&D Systems) according to the manufacturer's instructions.

Statistical analysis

Groups were compared with Prism software (GraphPad).

REFERENCES AND NOTES

1. M. G. Netea *et al*., Defining trained immunity and its role in health and disease. *Nat. Rev. Immunol.* **20**, 375–388 (2020). doi: 10.1038/s41577-020-0285-6; pmid: 32132681
2. S. Naik *et al*., Inflammatory memory sensitizes skin epithelial stem cells to tissue damage. *Nature* **550**, 475–480 (2017). doi: 10.1038/nature24271; pmid: 29045388
3. J. Ordovas-Montanes *et al*., Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649–654 (2018). doi: 10.1038/s41586-018-0449-8; pmid: 30135581
4. R. S. Moore, R. Kaletsky, C. T. Murphy, Piwi/PRG-1 argonaute and TGF-β mediate transgenerational learned pathogenic avoidance. *Cell* **177**, 1827–1841.e12 (2019). doi: 10.1016/j.cell.2019.05.024; pmid: 31178117
5. G. Tetreau, J. Dhinaut, B. Gourbal, Y. Moret, Trans-generational immune priming in invertebrates: Current knowledge and future prospects. *Front. Immunol.* **10**, 1938 (2019). doi: 10.3389/fimmu.2019.01938; pmid: 31475001
6. M. L. T. Berendsen *et al*., Maternal priming: Bacillus Calmette-Guérin (BCG) vaccine scarring in mothers enhances the survival of their child with a BCG vaccine scar. *J. Pediatric Infect. Dis. Soc.* **9**, 166–172 (2020). doi: 10.1093/jpids/piy142; pmid: 30715451
7. B. de Laval *et al*., C/EBPβ-dependent epigenetic memory induces trained immunity in hematopoietic stem cells. *Cell Stem Cell* **26**, 657–674.e8 (2020). doi: 10.1016/j.stem.2020.01.017; pmid: 32169166
8. S. P. Rosshart *et al*., Wild mouse gut microbiota promotes host fitness and improves disease resistance. *Cell* **171**, 1015–1028.e13 (2017). doi: 10.1016/j.cell.2017.09.016; pmid: 29056339
9. L. K. Beura *et al*., Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature* **532**, 512–516 (2016). doi: 10.1038/nature17655; pmid: 27096360
10. G. Mor, P. Aldo, A. B. Alvero, The unique immunological and microbial aspects of pregnancy. *Nat. Rev. Immunol.* **17**, 469–482 (2017). doi: 10.1038/nri.2017.64; pmid: 28627518
11. S. Niewiesk, Maternal antibodies: Clinical significance, mechanism of interference with immune responses, and possible vaccination strategies. *Front. Immunol.* **5**, 446 (2014). doi: 10.3389/fimmu.2014.00446; pmid: 25278941
12. G. Caballero-Flores *et al*., Maternal immunization confers protection to the offspring against an attaching and effacing pathogen through delivery of IgG in breast milk. *Cell Host Microbe* **25**, 313–323.e4 (2019). doi: 10.1016/j.chom.2018.12.015; pmid: 30686564
13. M. G. Darby *et al*., Pre-conception maternal helminth infection transfers via nursing long-lasting cellular immunity against helminths to offspring. *Sci. Adv.* **5**, eaav3058 (2019). doi: 10.1126/sciadv.aav3058; pmid: 31236458
14. E. Y. Hsiao, S. W. McBride, J. Chow, S. K. Mazmanian, P. H. Patterson, Modeling an autism risk factor in mice leads to permanent immune dysregulation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12776–12781 (2012). doi: 10.1073/pnas.1202556109; pmid: 22802640
15. R. L. Goldenberg, J. C. Hauth, W. W. Andrews, Intrauterine infection and preterm delivery. *N. Engl. J. Med.* **342**, 1500–1507 (2000). doi: 10.1056/NEJM200005183422007; pmid: 10816189
16. R. Medzhitov, D. S. Schneider, M. P. Soares, Disease tolerance as a defense strategy. *Science* **335**, 936–941 (2012). doi: 10.1126/science.1214935; pmid: 22363001
17. J. L. McCarville, J. S. Ayres, Disease tolerance: Concept and mechanisms. *Curr. Opin. Immunol.* **50**, 88–93 (2018). doi: 10.1016/j.coi.2017.12.003; pmid: 29253642
18. S. J. Han *et al*., White adipose tissue is a reservoir for memory T cells and promotes protective memory responses to infection. *Immunity* **47**, 1154–1168.e6 (2017). doi: 10.1016/j.immuni.2017.11.009; pmid: 29221731
19. M. W. McCoy, M. L. Marré, C. F. Lesser, J. Mecsas, The C-terminal tail of *Yersinia pseudotuberculosis* YopM is critical for interacting with RSK1 and for virulence. *Infect. Immun.* **78**, 2584–2598 (2010). doi: 10.1128/IAI.00141-10; pmid: 20368345
20. K. Miyawaki *et al*., CD41 marks the initial myelo-erythroid lineage specification in adult mouse hematopoiesis: Redefinition of murine common myeloid progenitor. *Stem Cells* **33**, 976–987 (2015). doi: 10.1002/stem.1906; pmid: 25446279
21. K. E. McGrath *et al*., Distinct sources of hematopoietic progenitors emerge before HSCs and provide functional blood cells in the mammalian embryo. *Cell Rep.* **11**, 1892–1904 (2015). doi: 10.1016/j.celrep.2015.05.036; pmid: 26095363
22. Y. Cong, T. Feng, K. Fujihashi, T. R. Schoeb, C. O. Elson, A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19256–19261 (2009). doi: 10.1073/pnas.0812681106; pmid: 19889972
23. T. W. Hand *et al*., Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses. *Science* **337**, 1553–1556 (2012). doi: 10.1126/science.1220961; pmid: 22923434
24. T. Sano *et al*., An IL-23R/IL-22 circuit regulates epithelial serum amyloid A to promote local effector T_H17 responses. *Cell* **163**, 381–393 (2015). doi: 10.1016/j.cell.2015.08.061; pmid: 26411290
25. M. J. Molloy *et al*., Intraluminal containment of commensal outgrowth in the gut during infection-induced dysbiosis. *Cell Host Microbe* **14**, 318–328 (2013). doi: 10.1016/j.chom.2013.08.003; pmid: 24034617
26. V. J. Carrión *et al*., Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science* **366**, 606–612 (2019). doi: 10.1126/science.aaw9285; pmid: 31672892
27. A. Stacy *et al*., Infection trains the host for microbiota-enhanced resistance to pathogens. *Cell* **184**, 615–627.e17 (2021). doi: 10.1016/j.cell.2020.12.011; pmid: 33453153
28. M. V. Zaretsky, J. M. Alexander, W. Byrd, R. E. Bawdon, Transfer of inflammatory cytokines across the placenta. *Obstet. Gynecol.* **103**, 546–550 (2004). doi: 10.1097/01.AOG.0000114980.40445.83; pmid: 14990420
29. S. C. Ganal-Vonarburg, M. W. Hornef, A. J. Macpherson, Microbial-host molecular exchange and its functional consequences in early mammalian life. *Science* **368**, 604–607 (2020). doi: 10.1126/science.aba0478; pmid: 32381716
30. J. Hoge *et al*., IL-6 controls the innate immune response against Listeria monocytogenes via classical IL-6 signaling. *J. Immunol.* **190**, 703–711 (2013). doi: 10.4049/jimmunol.1201044; pmid: 23241882
31. S. M. Chen, C. P. Lin, J. D. Tsai, Y. H. Chao, J. N. Sheu, The significance of serum and fecal levels of interleukin-6 and interleukin-8 in hospitalized children with acute rotavirus and norovirus gastroenteritis. *Pediatr. Neonatol.* **55**, 120–126 (2014). doi: 10.1016/j.pedneo.2013.05.008; pmid: 23899552
32. G. B. Choi *et al*., The maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring. *Science* **351**, 933–939 (2016). doi: 10.1126/science.aad0314; pmid: 26822608
33. J. R. Prins, N. Gomez-Lopez, S. A. Robertson, Interleukin-6 in pregnancy and gestational disorders. *J. Reprod. Immunol.* **95**, 1–14 (2012). doi: 10.1016/j.jri.2012.05.004; pmid: 22819759
34. S. Kim *et al*., Maternal gut bacteria promote neurodevelopmental abnormalities in mouse offspring. *Nature* **549**, 528–532 (2017). doi: 10.1038/nature23910; pmid: 28902840
35. Y. Lee *et al*., Induction and molecular signature of pathogenic TH17 cells. *Nat. Immunol.* **13**, 991–999 (2012). doi: 10.1038/ni.2416; pmid: 22961052
36. J. Y. Lee *et al*., Serum amyloid A proteins induce pathogenic T_H17 cells and promote inflammatory disease. *Cell* **180**, 79–91.e16 (2020). doi: 10.1016/j.cell.2019.11.026; pmid: 31866067
37. M. Murakami, D. Kamimura, T. Hirano, Pleiotropy and specificity: Insights from the interleukin 6 family of cytokines. *Immunity* **50**, 812–831 (2019). doi: 10.1016/j.immuni.2019.03.027; pmid: 30995501
38. J. Dahlgren, A. M. Samuelsson, T. Jansson, A. Holmäng, Interleukin-6 in the maternal circulation reaches the rat fetus in mid-gestation. *Pediatr. Res.* **60**, 147–151 (2006). doi: 10.1203/01.pdr.0000230026.74139.18; pmid: 16864694
39. L. Goetzl, T. Evans, J. Rivers, M. S. Suresh, E. Lieberman, Elevated maternal and fetal serum interleukin-6 levels are associated with epidural fever. *Am. J. Obstet. Gynecol.* **187**, 834–838 (2002). doi: 10.1067/mob.2002.127135; pmid: 12388959
40. D. Fawkner-Corbett *et al*., Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810–826.e23 (2021). doi: 10.1016/j.cell.2020.12.016; pmid: 33406409

41. M. P. Verzi *et al*., Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Dev. Cell* **19**, 713–726 (2010). doi: 10.1016/j.devcel.2010.10.006; pmid: 21074721
42. U. Jadhav *et al*., Dynamic reorganization of chromatin accessibility signatures during dedifferentiation of secretory precursors into Lgr5+ intestinal stem cells. *Cell Stem Cell* **21**, 65–77.e5 (2017). doi: 10.1016/j.stem.2017.05.001; pmid: 28648363
43. R. Francis *et al*., Gastrointestinal transcription factors drive lineage-specific developmental programs in organ specification and cancer. *Sci. Adv.* **5**, eaax8898 (2019). doi: 10.1126/sciadv.aax8898; pmid: 31844668
44. J. Guiu *et al*., Tracing the origin of adult intestinal stem cells. *Nature* **570**, 107–111 (2019). doi: 10.1038/s41586-019-1212-5; pmid: 31092921
45. A. M. Chin, D. R. Hill, M. Aurora, J. R. Spence, Morphogenesis and maturation of the embryonic and postnatal intestine. *Semin. Cell Dev. Biol.* **66**, 81–93 (2017). doi: 10.1016/j.semcdb.2017.01.011; pmid: 28161556
46. Z. Al Nabhani *et al*., A weaning reaction to microbiota is required for resistance to immunopathologies in the adult. *Immunity* **50**, 1276–1288.e5 (2019). doi: 10.1016/j.immuni.2019.02.014; pmid: 30902637
47. N. Barker *et al*., Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007). doi: 10.1038/nature06196; pmid: 17934449
48. A. L. Haber *et al*., A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017). doi: 10.1038/nature24489; pmid: 29144463
49. M. Biton *et al*., T helper cell cytokines modulate intestinal stem cell renewal and differentiation. *Cell* **175**, 1307–1320.e22 (2018). doi: 10.1016/j.cell.2018.10.008; pmid: 30392957
50. M. Koyama *et al*., MHC class II antigen presentation by the intestinal epithelium initiates graft-versus-host disease and is influenced by the microbiota. *Immunity* **51**, 885–898.e887 (2019). doi: 10.1016/j.immuni.2019.08.011; pmid: 31542340
51. E. M. Velazquez *et al*., Endogenous Enterobacteriaceae underlie variation in susceptibility to Salmonella infection. *Nat. Microbiol.* **4**, 1057–1064 (2019). doi: 10.1038/s41564-019-0407-8; pmid: 30911125
52. B. Chassaing *et al*., Fecal lipocalin 2, a sensitive and broadly dynamic non-invasive biomarker for intestinal inflammation. *PLOS ONE* **7**, e44328 (2012). doi: 10.1371/journal.pone.0044328; pmid: 22957064
53. M. J. McGeachy, S. J. McSorley, Microbial-induced T_H17: Superhero or supervillain? *J. Immunol.* **189**, 3285–3291 (2012). doi: 10.4049/jimmunol.1201834; pmid: 22997231
54. K. Geddes *et al*., Identification of an innate T helper type 17 response to intestinal bacterial pathogens. *Nat. Med.* **17**, 837–844 (2011). doi: 10.1038/nm.2391; pmid: 21666695
55. B. Chassaing, J. D. Aitken, M. Malleshappa, M. Vijay-Kumar, Dextran sulfate sodium (DSS)-induced colitis in mice. *Curr. Protoc. Immunol.* **104**, 15.25.1–15.25.14 (2014). doi: 10.1002/0471142735.im1525s104; pmid: 24510619
56. C. Asseman, S. Read, F. Powrie, Colitogenic Th1 cells are present in the antigen-experienced T cell pool in normal mice: Control by CD4+ regulatory T cells and IL-10. *J. Immunol.* **171**, 971–978 (2003). doi: 10.4049/jimmunol.171.2.971; pmid: 12847269
57. T. Krausgruber *et al*., T-bet is a key modulator of IL-23-driven pathogenic CD4(+) T cell responses in the intestine. *Nat. Commun.* **7**, 11627 (2016). doi: 10.1038/ncomms11627; pmid: 27193261
58. S. Beyaz *et al*., High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* **531**, 53–58 (2016). doi: 10.1038/nature17173; pmid: 26935695
59. L. M. Christian, K. Porter, Longitudinal changes in serum proinflammatory markers across pregnancy and postpartum: Effects of maternal body mass index. *Cytokine* **70**, 134–140 (2014). doi: 10.1016/j.cyto.2014.06.018; pmid: 25082648
60. P. Shooshtari *et al*., Correlation analysis of intracellular and secreted cytokines via the generalized integrated mean fluorescence intensity. *Cytometry A* **77**, 873–880 (2010). doi: 10.1002/cyto.a.20943; pmid: 20629196
61. E. Bolyen *et al*., Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019). doi: 10.1038/s41587-019-0209-9; pmid: 31341288
62. M. R. Corces *et al*., Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016). doi: 10.1038/ng.3646; pmid: 27526324
63. V. M. Link *et al*., Analysis of genetically diverse macrophages reveals local and domain-wide mechanisms that control transcription factor binding and function. *Cell* **173**, 1796–1809.e17 (2018). doi: 10.1016/j.cell.2018.04.018; pmid: 29779944
64. M. Stoeckius *et al*., Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018). doi: 10.1186/s13059-018-1603-1; pmid: 30567574

ACKNOWLEDGMENTS

We thank T. Hawley, J. Chen, S. Mistry, G. Koroleva, K. Beacht, E. Lewis, the NIAID animal facility, and the NIAID Microbiome Program gnotobiotic animal facility for technical support. We thank K. Mao, A. Gola, and the NIAID Biological Imaging Facility for providing support for confocal imaging. We thank all Belkaid laboratory members, especially S. Tamoutounour, P. Kulalert, M. Enamorado, and E. Ansaldo, for providing constructive feedback on this project and for critical reading of this manuscript. **Funding:** Y.B. is supported by the Division of Intramural Research of NIAID (grants ZIA-AI001115 and ZIA-AI001132). H.A.C. is supported by the Division of Intramural Research of the NIMH (grant ZIA-MH002784). M.S.L. is supported by the Division of Intramural Research of the NIAID (grant ZIA-AI001175). A.I.L. is supported by the Human Frontier Science Program (grant LT000191/2018). A.S. is supported by the National Institute of General Medical Sciences (NIGMS) Postdoctoral Research Associate (PRAT) fellowship program (grant 1FI2GM128736). **Author contributions**: A.I.L. and Y.B. designed the study experiments and wrote the manuscript. A.I.L. performed experiments and analyzed the data. T.M., T.K.F., and D.S.L.-J. participated in performing experiments. V.M.L. analyzed RNA-seq and ATAC-seq data. A.S provided expertise in *Salmonella* Typhimurium infection and analyzed 16*S* data. H.-Y.S. provided expertise in ATAC-seq. O.J.H. provided intellectual expertise. J.V.D. and M.S.L. provided expertise in *C. albicans* infection. R.-M.K. and H.A.C. performed and interpreted data from behavior studies. S.-J.H. provided intellectual expertise, participated in experiments, and helped to interpret experimental results. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** 16*S* sequencing, ATAC-seq, RNA-seq, and scRNA-seq data were deposited into the NCBI Sequence Read Archive (accession numbers PRJNA666968, GSE158950, and GSE159149). All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/eabf3002/suppl/DC1
Figs. S1 to S9
Table S1
MDAR Reproducibility Checklist

18 October 2020; accepted 7 July 2021
10.1126/science.abf3002

RESEARCH ARTICLE SUMMARY

NEURODEVELOPMENT

Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells

Ioannis Sarropoulos*†, Mari Sepp*†, Robert Frömel, Kevin Leiss, Nils Trost, Evgeny Leushkin, Konstantin Okonechnikov, Piyush Joshi, Peter Giere, Lena M. Kutscher, Margarida Cardoso-Moreira, Stefan M. Pfister*‡, Henrik Kaessmann*‡

INTRODUCTION: The cerebellum contributes to many complex brain functions, including motor control, language, and memory. During development, distinct neural cells are generated at cerebellar germinal zones in a spatiotemporally restricted manner. Cis-regulatory elements (CREs), such as enhancers and promoters, and the transcription factors that bind to them are central to cell fate specification and differentiation. Although most CREs undergo rapid turnover during evolution, a few are conserved across vertebrates.

RATIONALE: Bulk measurements of CRE activity have provided insights into gene regulation in the cerebellum, as well as into the evolutionary dynamics of CREs during organ development. However, they lack the cellular resolution required to assess cell-type differences in regulatory constraint and unravel the regulatory programs associated with the specification and differentiation of cell types.

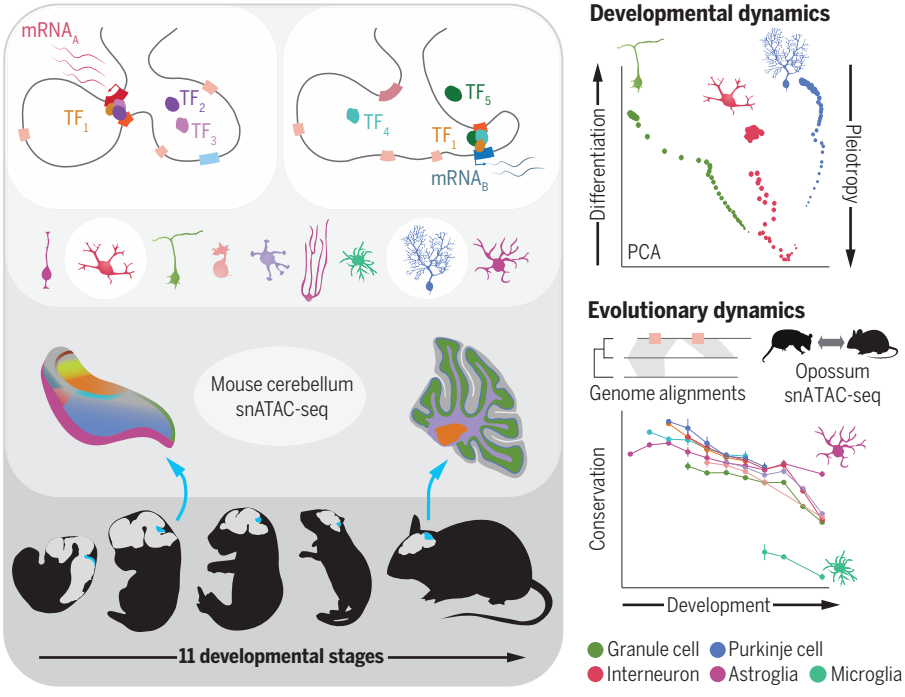
RESULTS: Here, we generated a single-cell atlas of gene regulation in the mouse cerebellum spanning 11 developmental stages, from the beginning of neurogenesis to adulthood. By acquiring snATAC-seq (single-nucleus assay for transposase accessible chromatin using sequencing) profiles for ~90,000 cells, we mapped all major cerebellar cell types and identified candidate CREs.

Characterization of CRE activity across the cerebellum development highlights the cell- and time-specificity of gene regulation. Many of the differentially accessible CREs are specific to a single cell type and state, but we also identified a fraction of CREs with pleiotropic (shared) activity. At early developmental stages, temporal changes in CRE activity are shared between progenitor cells from different germinal zones, supporting a model of cell fate induction through common temporal cues. Pleiotropic CREs in major cerebellar neuron types (granule cells, Purkinje cells, and inhibitory interneurons) are more active at early differentiation states, and the regulatory programs gradually diverge as differentiation proceeds.

Based on comparisons to vertebrate genomes, we observed a decrease in CRE sequence conservation during development for all cerebellar cell types, a pattern that is largely explained by differentiation as well as by additional temporal differences between cells from matched differentiation states. Across cell types, differences in regulatory conservation are most pronounced in the adult, where microglia—the immune cells of the brain—show the fastest evolutionary turnover. By contrast, mature astrocytes harbor the most conserved intergenic CREs, not only in the cerebellum but also across a wide range of cell types in adult mouse organs.

To evaluate the conservation of CRE activity, we acquired snATAC-seq profiles for ~20,000 cerebellar cells from the gray short-tailed opossum, a marsupial separated from mouse by ~160 million years of evolution. Our comparative analysis of CRE activity in the two therian species reinforced our sequence-based conclusions regarding differences in CRE constraint across cell types and developmental stages and also revealed that despite the overall high turnover of CREs, radical repurposing of spatiotemporal CRE activity is rare, at least between cell types in the same tissue.

CONCLUSION: This study reveals extensive temporal differences in CRE activity across cerebellar cell types and a shared decrease in CRE conservation during development and differentiation. Given that the cerebellum has been successfully used as a model system to study cell fate specification, neurogenesis, and other developmental processes, we expect that our observations regarding the developmental and evolutionary dynamics of regulatory elements, and their interplay, are also applicable to mammalian organs in general. ■



Cis-regulatory elements in cerebellar cells. snATAC-seq delineates cell- and time-specific CRE activity in the developing mouse cerebellum (left). The chromatin accessibility profiles of cerebellar neuron types gradually diverge during differentiation as the activity of pleiotropic (shared) CREs decreases (top right). The evolutionary conservation of CRE sequences in vertebrates and activity in therian mammals decreases across development and differs between cell types (bottom right). mRNA, messenger RNA; PCA, principal components analysis; TF, transcription factor.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: h.kaessmann@zmbh.uni-heidelberg.de (H.K.); s.pfister@dkfz.de (S.M.P.); i.sarropoulos@zmbh.uni-heidelberg.de (I.S.); m.sepp@zmbh.uni-heidelberg.de (M.S.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as I. Sarropoulos *et al.*, *Science* **373**, eabg4696 (2021). DOI: 10.1126/science.abg4696

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg4696

RESEARCH ARTICLE

NEURODEVELOPMENT

Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells

Ioannis Sarropoulos[1]*†, Mari Sepp[1]*†, Robert Frömel[1]‡§, Kevin Leiss[1], Nils Trost[1], Evgeny Leushkin[1], Konstantin Okonechnikov[2], Piyush Joshi[2], Peter Giere[3], Lena M. Kutscher[4], Margarida Cardoso-Moreira[1,5], Stefan M. Pfister[2,6]*¶, Henrik Kaessmann[1]*¶

Organ development is orchestrated by cell- and time-specific gene regulatory networks. In this study, we investigated the regulatory basis of mouse cerebellum development from early neurogenesis to adulthood. By acquiring snATAC-seq (single-nucleus assay for transposase accessible chromatin using sequencing) profiles for ~90,000 cells spanning 11 stages, we mapped cerebellar cell types and identified candidate cis-regulatory elements (CREs). We detected extensive spatiotemporal heterogeneity among progenitor cells and a gradual divergence in the regulatory programs of cerebellar neurons during differentiation. Comparisons to vertebrate genomes and snATAC-seq profiles for ~20,000 cerebellar cells from the marsupial opossum revealed a shared decrease in CRE conservation during development and differentiation as well as differences in constraint between cell types. Our work delineates the developmental and evolutionary dynamics of gene regulation in cerebellar cells and provides insights into mammalian organ development.

The cerebellum, known for its role in motor control, also contributes to other complex functions, such as language and memory (*1*–*3*). During development, two germinal zones—the ventricular zone and the rhombic lip—give rise to distinct neuronal and glial populations (*2*, *3*). Cell-type specification and differentiation processes are controlled by cis-regulatory elements (CREs), such as enhancers and promoters, and the transcription factors (TFs) that bind to them (*4*, *5*). Although most CREs evolve rapidly (*6*–*8*), some, enriched near developmental genes, are conserved across vertebrates (*9*, *10*). Studies of bulk CRE activity have explored gene regulation in the hindbrain and postnatal cerebellum (*11*, *12*), as well as CRE evolutionary dynamics during organ development (*13*, *14*). Single-cell approaches have been used to study gene regulation in adult mouse organs (*15*), including the brain (*16*), and in a limited number of cells or stages for the developing mouse (*17*, *18*) and human brain (*19*, *20*), but a comprehensive characterization of the entire development of a mammalian organ has been lacking. In this study, we used single-nucleus assay for transposase accessible chromatin using sequencing (snATAC-seq) (*21*) to profile the landscape of chromatin accessibility—a proxy for CRE activity—of ~90,000 cells across pre- and postnatal development of the mouse cerebellum and ~20,000 cerebellar cells from the marsupial opossum (https://apps.kaessmannlab.org/mouse_cereb_atac/). We leveraged this dataset to assess the developmental and evolutionary dynamics of gene regulation in the cerebellum.

A snATAC-seq atlas of mouse cerebellum development

We generated snATAC-seq data for male and female mouse cerebella from 11 stages ranging from early neurogenesis [embryonic day 10 (E10)] to adulthood [postnatal day 63 (P63); Fig. 1A, fig. S1, and table S1]. After applying quality controls (*22*), we acquired single-cell chromatin accessibility profiles for 91,922 cells with a median of 20,558 fragments per cell (Fig. 1B; fig. S2, A to D; and table S2). We projected cells into a low-dimensional embedding (excluding sex chromosomes) and identified clusters (*22*). Biological replicates show high similarity in this embedding (Fig. 1B), distribute evenly across clusters (fig. S2E), and have the highest correlations in the accessibility of autosomal CREs at the library level, followed by correlations between samples from adjacent developmental stages (fig. S2F).

Using iterative clustering (fig. S3) and approximating gene expression by the aggregated accessibility across a gene's regulatory region (gene score) (*22*), we identified 12 broad cell types and 42 subtypes and cell states (Fig. 1, C to E; fig. S4; and table S2). To assess the use of gene scores as a proxy for gene expression and the quality of our cell-type annotation, we reprocessed the data by developmental stage (fig. S5) and integrated with a single-cell RNA sequencing (scRNA-seq) atlas of the developing mouse hindbrain and cerebellum (*23*), transferring labels and imputing RNA expression for cells in our snATAC-seq atlas. Despite differences in sampling, we observed good correlations between gene scores and imputed RNA expression values (fig. S6A), and a concordance between cell-type labels (fig. S6, B and C). Discrepancies between the two annotations are explained by some scRNA-seq clusters containing mixtures of cell types (fig. S6, D and E).

The earliest developmental stages in our dataset (E10 and E11) are dominated by neural progenitors (key marker genes: *Notch1*, *Cyp26b1*; Fig. 1, C to E). We also detected midbrain-originating cells (*Isl1*) and parabrachial neurons (*Lmx1b*), which later migrate out of the cerebellum (*24*, *25*), as well as differentiating γ-aminobutyric acid–producing (GABAergic) deep nuclei neurons (*Zfhx3*, *Sox14*) and glutamatergic neurons (*Meis2*, *Neurod6*) of the nuclear transitory zone. From E12 onward, the latter could be further resolved into posterior (*Lmx1a*) and ventral (*Lmo3*) populations of glutamatergic deep nuclei neurons and into anteriorly located isthmic nuclei neurons (*Pax5*) (*24*). The same developmental window is marked by the generation of Purkinje cells (*Skor2*, *Foxp2*), which are the most abundant neuron type until E15. During the E10 to E15 period, the relative abundance of progenitor cells (excluding granule cell progenitors) decreases from 83 to 17% (Fig. 1D).

The next stages (E17 to P4) are marked by the emergence of GABAergic interneurons (*Pax2*) and unipolar brush cells (UBCs; *Lmx1a*, *Eomes*) and the rapid expansion of the proliferating (*Atoh1*, *Gli2*) and differentiating (*Neurod1*, *Grin2b*) granule cell populations (Fig. 1, C to E). Small numbers of microglia (*Ccr1*, *Hexb*) and oligodendrocytes (*Sox10*), most of which are of extracerebellar origin (*2*), are detectable from E17 onward. The last two stages (P14 and P63) are dominated by mature granule cells (*Etv1*, *Cbln3*), with additional neuronal and glial populations being traceable, including mature astroglia such as Bergmann glia (*Gdf10*) and parenchymal astrocytes (*Aqp4*, *Scl6a11*). Throughout this study, we use the term astroglia to refer to cells transitioning along the lineage of neuroepithelial progenitors, radial glial progenitors, and mature astrocytes (Fig. 1, C to E), in agreement with their overlapping

[1]Center for Molecular Biology of Heidelberg University (ZMBH), DKFZ-ZMBH Alliance, D-69120 Heidelberg, Germany. [2]Hopp Children's Cancer Center (KiTZ) Heidelberg, Division of Pediatric Neurooncology, German Cancer Consortium (DKTK), and German Cancer Research Center (DKFZ), D-69120 Heidelberg, Germany. [3]Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany. [4]Hopp Children's Cancer Center (KiTZ) Heidelberg, Developmental Origins of Pediatric Cancer Group, German Cancer Research Center (DKFZ), D-69120 Heidelberg, Germany. [5]Evolutionary Developmental Biology Laboratory, Francis Crick Institute, London NW1 1AT, UK. [6]Department of Pediatric Hematology and Oncology, Heidelberg University Hospital, Heidelberg, Germany.
***Corresponding author. Email: h.kaessmann@zmbh.uni-heidelberg.de (H.K.); s.pfister@dkfz.de (S.M.P.); i.sarropoulos@zmbh.uni-heidelberg.de (I.S.); m.sepp@zmbh.uni-heidelberg.de (M.S.)**
†These authors contributed equally to this work.
‡Present address: Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona 08003, Spain.
§Present address: Universitat Pompeu Fabra, Barcelona, Spain.
¶These authors contributed equally to this work.

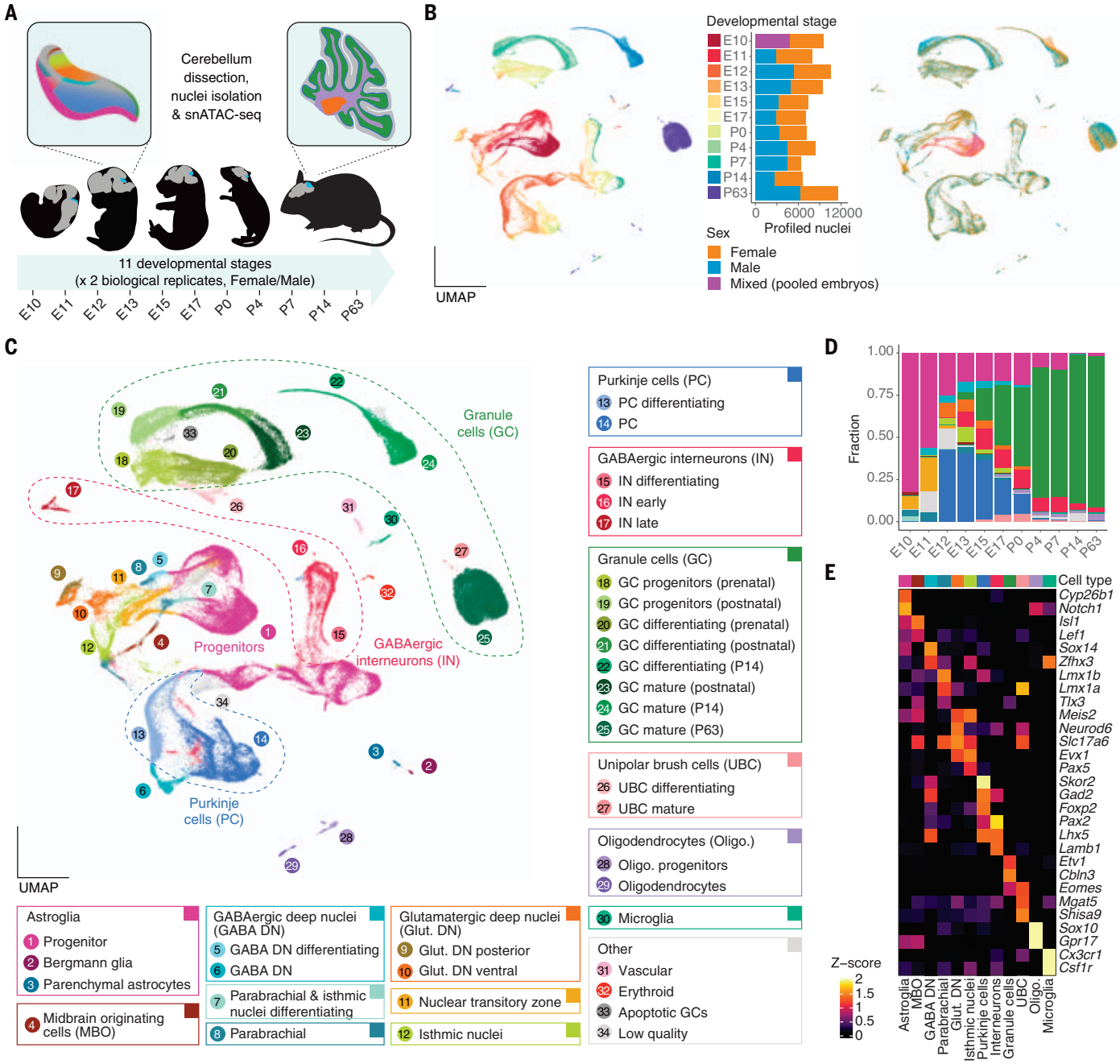


Fig. 1. A snATAC-seq atlas of mouse cerebellum development. (**A**) Schematic overview of the dataset. Representative mouse silhouettes are shown for E11, E13, E17, P4, and P63 (brain in gray, cerebellum in cyan). The insets show the location of selected cell types in the cerebellum [colors are as in (C)]. (**B** and **C**) Uniform manifold approximation and projection (UMAP) of 91,922 cells colored by developmental stage [(B), left] or sex [(B), right], or cell type and state (C). Barplots in (B) show the number of profiled cells per stage and sex (with each sex corresponding to one sample). In (C), cell states or subtypes (numbered circles) are grouped into broad cell types (rectangles). Further progenitor and interneuron subtypes are shown in Fig. 3 and figs. S11 and S13. (**D**) Proportions of broad cell types across developmental stages. (**E**) Activity scores of genes used for the annotation of broad cell types (*z*-score, capped to 0 to 2). Broad cell-type colors for (D) and (E) are as in (C). Marker genes across subtypes and states are shown in fig. S4.

functions and molecular features (*26*). Overall, the developmental dynamics of cerebellar cell abundances observed in this study resemble those from scRNA-seq atlases (fig. S6F) (*23*, *27*). However, some cell types, such as early-born neurons (GABAergic and glutamatergic deep nuclei, isthmic nuclei, and parabrachial neurons) are better resolved in our snATAC-seq atlas.

The cis-regulatory landscape of cerebellar cell types

To characterize the regulatory profiles of cerebellar cell types, we used a cluster-specific and replicate-aware peak calling approach (*22*). We identified 261,643 high-confidence putative CREs (Fig. 2A and fig. S7, A and B), most of which are intronic (51%) or intergenic (26%) (herein collectively referred to as distal). These putative CREs cover ~80% of enhancers active during hindbrain development (*11*),

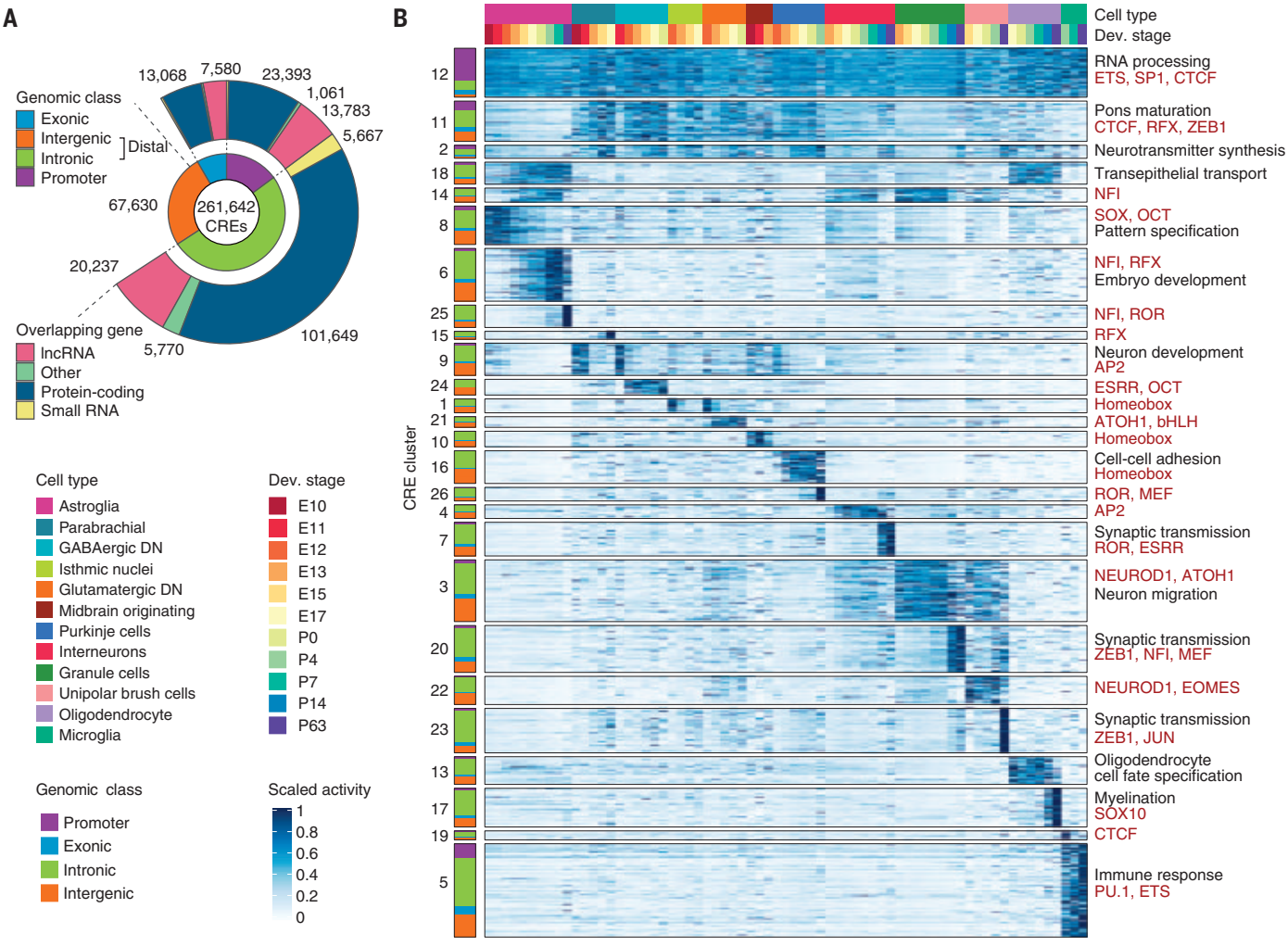


Fig. 2. CRE activity in cerebellum development. (**A**) Genomic features of 261,642 putative CREs. The inner circle shows genomic class; the outer circle shows the biotype of the overlapping gene. (**B**) Clusters of CRE activity across cell types and developmental stages. CREs are grouped by activity cluster (*k*-means followed by hierarchical clustering) and genomic class (left). CRE clusters are arranged in decreasing order of pleiotropy (here as mean activity across rows) and then by cell type and developmental stage with maximum activity. Shown on the right are representative enrichments [Benjamini-Hochberg (BH)–adjusted $P < 0.05$; hypergeometric test] for biological processes of adjacent genes (black) and motifs for TFs or TF families (red). 50,000 CREs confidently assigned to their cluster were chosen randomly for visualization.

including 97% of experimentally validated elements in the E11.5 hindbrain (*28*) and 82% of enhancer RNAs (eRNAs) transcribed in the developing cerebellum (*29*), a substantial enrichment compared with heterochromatin or enhancers active in non-neuronal tissues (fig. S7, C to G). Comparisons with a single-cell ATAC-seq atlas of adult mouse organs (*15*) revealed the highest activity of the putative CREs in cells from the cerebellum (fig. S7H).

We assigned 32,792 distal CREs to 5766 putative target genes based on correlations with promoter accessibility and gene scores (fig. S8A and table S3). The CRE-gene pairs share more often the same topologically associating domain (TAD) in neural progenitors (*30*), are enriched for enhancer activity in the hindbrain (*11*, *28*), and show higher correlations between promoter transcripts and eRNAs in the developing cerebellum (*29*) (fig. S8, B to E). Although most genes were associated with a single distal CRE, 1001 protein-coding genes were assigned to 10 or more elements (fig. S8F). These genes are enriched for developmental processes, including cell fate commitment and neuron differentiation, as well as mature neuron functions like synapse organization and signaling (fig. S8G).

To describe patterns of CRE activity during cerebellum development, we used an iterative clustering procedure (fig. S9, A and B). Most of the 26 CRE clusters are cell type– and time-specific, whereas one shows constitutive activity (cluster 12; Fig. 2B). Most (67%) of these constitutive CREs correspond to promoters (fig. S9C), and among the remaining distal CREs, more than 50% contain a CTCF motif (fig. S9D), suggesting architectural roles, in accord with observations in humans (*20*). We further observed groups of CREs active in multiple early-born neuron types (clusters 2 and 11), glial populations (cluster 18), and late-born cell types (cluster 14), affirming that a sizeable fraction of CREs have pleiotropic activity (*31*, *32*).

CREs in cell type–specific clusters are close to genes associated with relevant gene ontology terms (e.g., myelination for oligodendrocyte-specific CREs) and are enriched for motifs of TFs known to be active in the respective cell types (e.g., ATOH1 for granule cells, SOX family TFs for progenitors, PU.1 for microglia; Fig. 2B and fig. S9, E and F). We refined this analysis by identifying 50,815 differentially accessible CREs with significantly higher activity in a cell type and state (fig. S10A). Some of the distinct states we discovered distinguish cells from the same cell type based on the developmental stage (e.g., pre- and postnatal granule cell progenitors, P14 and P63 mature granule

cells). The high similarity between biological replicates (fig. S2, E and F) argues against technical confounders as a source of this separation. Indeed, we detected 3988 differentially accessible CREs between pre- and postnatal granule cell progenitors (fig. S10, B and C). CREs with increasing accessibility after birth are enriched for motifs of NFI, known to specify late-born neurons (*33*), and GLI2, consistent with the sonic hedgehog (SHH)–mediated expansion of the granule cell progenitor pool after birth (*2*). CREs with decreasing accessibility are enriched for motifs marking embryonic progenitors (SOX, MEIS) (*2*). Altogether, our analyses provide a comprehensive characterization of CRE activity during cerebellum development and highlight the cell and time specificity of gene regulation.

Progenitor heterogeneity reflects cell fate decisions

Cerebellum development relies on a spatially and temporally restricted pattern of cell-type specification. GABAergic deep nuclei neurons, Purkinje cells, interneurons, and astrocytes are sequentially derived from progenitors in the ventricular zone, whereas the emergence of glutamatergic deep nuclei neurons from rhombic lip progenitors is followed by the generation of granule cells and UBCs (*2*). We asked whether this mode of cell fate specification was associated with heterogeneity among cerebellar progenitors. We subclustered cells from the astroglia lineage (progenitors and astrocytes) and identified progenitors from all germinal regions throughout cerebellum development, although without sharp boundaries within a given stage (Fig. 3, A to C; and fig. S11, A to C). Early progenitors (E10 to E12) include isthmic (*En1*, *Pax5*), ventricular zone (*Dll1*, *Ptf1a* from E11), and rhombic lip (*Cdon*, *Atoh1*) populations, as well as progenitors with no apparent commitment toward a cell fate (Fig. 3, A to C; and fig. S11, A to C), in accord with previous reports (*34*). We additionally identified a *Gsx1*$^+$ population that we traced to the anterior ventricular zone (fig. S11, A and B) (*35*). E13 to E15 are marked by the appearance of two late progenitor populations (fig. S11C) that broadly correspond to the previously described bipotent (*Gsx1*, *Wnt8b*) and gliogenic (*Slc1a3*, *Grm3*) progenitors (*2*, *36*). Previous studies suggest that the bipotent progenitors generate interneurons and white matter astrocytes, whereas the gliogenic progenitors give rise to Bergmann glia and granule cell layer astrocytes (*2*, *36*). Consistently, we detected two groups of differentiating parenchymal astrocytes at perinatal stages (E15 to P7): white matter (*Slc6a11*, *Olig2*, *Kcnd2*) and granule cell layer (*Aqp4*, *Tekt5*) astroblasts, with higher similarity in CRE activity to bipotent and gliogenic progenitors, respectively (Fig. 3, A to C; and fig. S11, A to D). We then investigated the relationship of the two *Gsx1*$^+$ populations: early anterior ventricular zone progenitors and late bipotent progenitors. Besides additional shared marker genes (*Wnt8b*, *Ndnf*, *Robo1*), E10 to E12 anterior ventricular zone progenitors show the highest similarity in CRE activity with E13 to E15 bipotent progenitors across stage-matched progenitor types (Fig. 3D and fig. S11, E and F). Previously, *Gsx1*$^+$ progenitors were shown to arise from a temporal transition of Purkinje cell–producing ventricular zone progenitors (*35*). Our data support the existence of an additional, molecularly distinct, anterior population of early progenitors that is already primed to acquire bipotent progenitor identity.

Despite our dense sampling and ability to identify the same germinal regions across consecutive stages, often based on the same marker genes (e.g., E10 to E12 rhombic lip; fig. S11A), progenitor cells primarily cluster by developmental stage and not by germinal zone (Fig. 3, A to C, and E), suggesting strong changes in their chromatin accessibility profiles during development. These temporal differences are strongest in early embryonic stages (E10 to E12), a period coinciding with the sequential generation of distinct neuronal cell types (Figs. 1D and 3E). Unbiased clustering of CRE activity across progenitor groups and developmental stages revealed many temporally dynamic CREs that are shared between germinal zones (Fig. 3F), arguing for common factors in control of temporal transitions in progenitor competence (*3*, *33*). These include a shift from CREs enriched for nuclear receptor and SOX motifs and associated with chromatin silencing genes (E10 to E12; clusters 2 and 3) toward CREs enriched for NFI motifs and adjacent to genes involved in developmental processes (E12 to E13; cluster 8), signaling, and cell adhesion (E13 to P7; cluster 5). From E15 onward, coinciding with the protracted generation of interneurons and astrocytes, temporal differences become smaller and progenitors from adjacent stages group by germinal region (gliogenic and bipotent; Fig. 3E).

To assess whether these temporal differences in chromatin accessibility also lead to changes in gene expression, we systematically identified genes with higher variance across developmental stages than between germinal zones (fig. S12, A and B). Reassuringly, marker genes such as *Gsx1* and *Gdf10* show higher variance across germinal zones. However, most of the 2000 highly variant genes in E10 to E13 show higher variance across developmental stages, whereas variance by progenitor type is higher in E15 to P0 (Fig. 3G), in agreement with our clustering analysis (Fig. 3E). Temporally variant genes (310 in E10 to E13 versus 61 in E15 to P0; table S4) include the pluripotency factor *Lin28a*, which is expressed in E10 progenitors across all germinal zones, and *Nfix* (Fig. 3H), which increases in expression during development, in line with the enrichment of NFI motifs in late CRE clusters (Fig. 3F). We next clustered temporally variant genes based on their activity profiles (decreasing or increasing; fig. S12, C and D) and examined their expression in progenitor cells in a published scRNA-seq dataset (*23*). We observed significant temporal differences in the direction predicted by the snATAC-seq data (fig. S12, E and F). Thus, although spatiotemporal heterogeneity among progenitors is prominent throughout cerebellum development, early stages are characterized by stronger temporal differences in CRE activity, which are often shared between germinal zones and lead to corresponding gene expression changes.

Regulatory activity during neuron differentiation

We next characterized the gene regulatory programs underlying the differentiation of the three most abundant neuron types in our dataset: granule cells, Purkinje cells, and interneurons. For each neuron type, we aligned cells from different developmental stages along their differentiation trajectories, which we modeled using diffusion pseudotime (fig. S13, A to C). The observed differentiation dynamics ranged from stage-restricted (Purkinje cells, E12 to E13) to protracted (granule cells, interneurons) (Fig. 4, A and B; and fig. S13D), consistent with the several-weeks-long generation of granule cells and interneurons (*2*). Whereas granule cells form a homogeneous differentiating population (Fig. 4B), interneurons are stratified into distinct temporally specified subtypes (fig. S13, E and F): early-born interneurons (*Zfhx4*, *Slit2*) detected at E13 to E15, mid-born Golgi cells (*Chrm2*), Purkinje layer interneurons (*Nxph1*, *Klhl1*) prevalent at E17 to P7, and late-born molecular layer interneurons of type 1 (*Sorcs3*, *Grm8*) and 2 (MLI2; *Nxph1*, *Pvalb*), which are abundant at P14 to P63 (*2*, *37*).

We used the modeled differentiation trajectories to characterize the gene regulatory dynamics in each neuron type. We identified TFs with dynamic activity during differentiation and correlated the accessibility of the respective binding motifs with their gene scores to classify TFs into putative activators and repressors (fig. S14, A to C; and table S5). These candidate TFs include known regulators, such as BARHL1 and ETV1 in granule cells and OLIG2 and FOXP2 in Purkinje cells (Fig. 4C).

Across all three neuron types, we observed a gradual transition from CREs proximal to genes associated with embryonic development and cell fate commitment toward genes with roles in neuron differentiation, migration, and axon guidance, and eventually the formation of the synapse, neurotransmitter secretion, and ion transport (Fig. 4D and fig. S14, D and E). In support of this convergence in biological

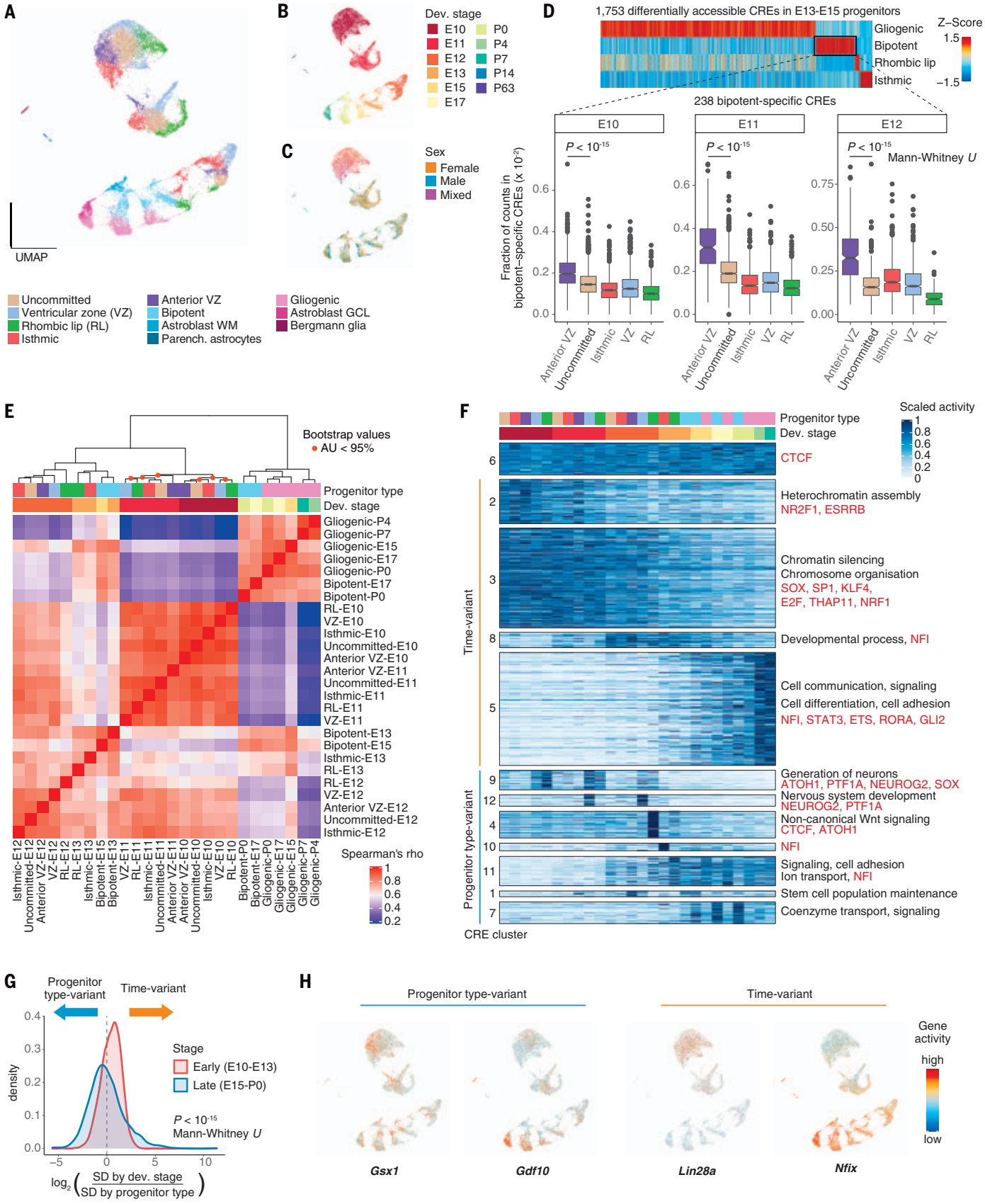


Fig. 3. Spatiotemporal heterogeneity in cerebellar progenitor populations. (**A** to **C**) UMAP projections of 21,830 astroglia cells (progenitors and astrocytes) colored by subtype (A), developmental stage (B), and sex (C). GCL, granule cell layer; WM, white matter. (**D**) Comparison of bipotent progenitors to earlier

populations. Activity profiles (*z*-score) of progenitor type–specific CREs in E13 to E15 are shown at the top. The fraction of fragments per cell in CREs specific to bipotent progenitors across progenitor types and developmental stages is shown at the bottom. Boxes represent the interquartile range, and whiskers extend to extreme values within 1.5 times the interquartile range from the box. (**E**) Hierarchical clustering based on Spearman's correlation coefficients in CRE accessibility across progenitor types and developmental stages. Orange dots indicate nodes with approximately unbiased (AU) probability values < 95%. (**F**) Clusters of CRE activity across progenitor types and developmental stages. CREs are grouped by activity cluster (*k*-means followed by hierarchical clustering). Representative enrichments (BH-adjusted $P < 0.05$; hypergeometric test) for biological processes of adjacent genes (black) and motifs for TFs or TF families (red) are shown on the right. 25,000 CREs confidently assigned to their cluster were chosen randomly for visualization. (**G**) Density distributions for the $\log_2$-ratio of gene score standard deviation (SD) across developmental stages and progenitor types for the 2000 genes with the highest variance in early and late progenitor populations. (**H**) Gene scores (capped at 10th and 99th quantiles and $\log_{10}$ transformed) for genes with high variance across progenitor types (left) or developmental stages (right).

processes, 43% of protein-coding genes with dynamic activity across pseudotime are shared by at least two neuron types (fig. S14F), in accordance with observations in the developing neocortex (*38*). By contrast, only 20% of dynamic CREs are shared across neuron types (i.e., are pleiotropic), further highlighting the cell type–specificity of gene regulation compared with gene expression (fig. S14G). Pleiotropic CREs show similar activity profiles across neuron types (Fig. 4E; early-early: $P < 10^{-15}$, late-late: $P < 10^{-15}$, early-late: $P > 0.99$, hypergeometric test) and, with the exception of interneurons, are more common in early stages of neuron differentiation (Fig. 4, F and G; and fig. S14, H and I). The higher similarity of the early stages is supported by a principal components analysis in which the chromatin accessibility profiles of the three neuron types gradually diverge during differentiation (Fig. 4H).

Evolutionary dynamics of CREs in cerebellum development

Motivated by the temporal dynamics in CRE activity during development (Fig. 2), cell-type specification (Fig. 3), and differentiation (Fig. 4), we next investigated whether such differences are also reflected in the evolutionary histories of CREs. A decrease in gene regulatory (*13*) and expression (*39*, *40*) conservation as development progresses has been reported for whole organs. Given the profound changes of cell-type abundances during development, it remains unclear whether whole-organ patterns emerge from the parallel decline of conservation across cell types or instead reflect the higher conservation of cell types that are more abundant in earlier stages, such as progenitors and early-born cells. Hence, we sought to test for differences in the evolutionary dynamics of CREs both across cell types and throughout development.

We assessed the selective (functional) constraints on CRE sequences using estimates of evolutionary conservation (phastCons scores) (*41*) and inferred their minimum age (i.e., when they first appeared) using syntenic sequence alignments between mouse and 16 other vertebrates at various phylogenetic distances (fig. S15 and table S6). We assessed these conservation metrics for intergenic CREs at the single-cell level, enabling comparisons between cell types and developmental stages. Across all cell types, both sequence constraints and the predicted ages of intergenic CREs decreased significantly during development (Fig. 5A and fig. S16A). Thus, previous observations for whole organs (*13*) are mostly explained by decreasing conservation within cell types rather than changes in the relative abundance of cell types with pronounced differences in evolutionary constraints. Ancient CREs, shared across or even beyond mammals, show higher activity during embryonic development, when cell types begin their differentiation (Fig. 5B). As cell types mature and activate their terminal effector genes, elements specific to eutherians and murid rodents (Muridae) gradually increase their activity, potentially contributing to species- or lineage-specific phenotypic innovations of ancestral cell types (Fig. 5B). Consistent with these observations, TF genes, which are central to cell-type identity (*42*), are associated with older and more constrained CREs (Fig. 5C and fig. S16B).

Given the strong developmental signals in our dataset, we next assessed the contribution of cell-type differentiation versus additional temporal differences, such as intrinsic patterning signals (*33*, *38*) or interactions with morphogens and ligands (*4*), to the temporal decline in conservation. We focused on granule cells because these cells have a protracted differentiation trajectory (E13 to P14) and, despite the different states observed in our data, are not known to form distinct temporally specified subtypes (*2*). By stratifying granule cells based on their differentiation state (pseudotime) and developmental stage, we observed that both factors contribute to the decrease of intergenic CRE conservation during development, with cells at early differentiation states and developmental stages showing the most conserved regulatory program (Fig. 5D and fig. S16, C and D). We validated this observation in a pseudotime-independent framework by focusing on prenatal granule cell progenitors, which formed a single cluster without any alignment across developmental stages (fig. S16E). Despite the overall high similarity of these cells, we identified CREs with significant temporal activity changes and clustered them into two groups that decrease (7527) or increase (11,972) in accessibility across prenatal stages (fig. S16E). CREs with decreasing accessibility are more conserved and are enriched for motifs recognized by pluripotency TFs (e.g., SOX2), whereas CREs with increasing accessibility are enriched for motifs of TFs active in granule cell progenitors, including ATOH1 and GLI2 (fig. S16, F and G). Using published scRNA-seq data (*23*), we observed similar developmental patterns for genes adjacent to temporally dynamic CREs (fig. S16H), suggesting that these differences in CRE accessibility affect gene expression.

Both CRE pleiotropy (Fig. 4, F to H) and sequence conservation (Fig. 5A) decrease during differentiation. Given that pleiotropy imposes constraints on the evolution of regulatory sequences (*31*, *43*), we asked whether it could explain the decrease in CRE conservation. We found that pleiotropic CREs are significantly more conserved than those dynamic in a single neuron type and that these differences are larger than those between early- and late-acting CREs with similar pleiotropy status (Fig. 5E). Consistently, CREs with decreasing accessibility across developmental stages in granule cell progenitors are more conserved (fig. S16F) and active in other types of progenitor and early differentiating cells, whereas CREs with increasing accessibility are specific to granule cell progenitors (fig. S16I). Thus, as CRE pleiotropy decreases during development and differentiation, constraint on CRE sequences relaxes, consistent with previous observations for gene expression (*39*, *40*).

Differences in CRE constraint across cell types

We also observed differences in the conservation of intergenic CREs between cell types. CREs active in microglia show a faster evolutionary turnover compared with all other cerebellar cell types (Fig. 5A and fig. S16A), consistent with the rapid divergence in gene expression and morphology of microglia (*44*). In agreement with their younger age and reduced sequence constraint, CREs active in microglia are also enriched for genomic repeats, in particular for recently expanded transposable elements in rodents, such as short interspersed nuclear elements (SINEs) B1, B2, and B4; endogenous retrovirus sequences

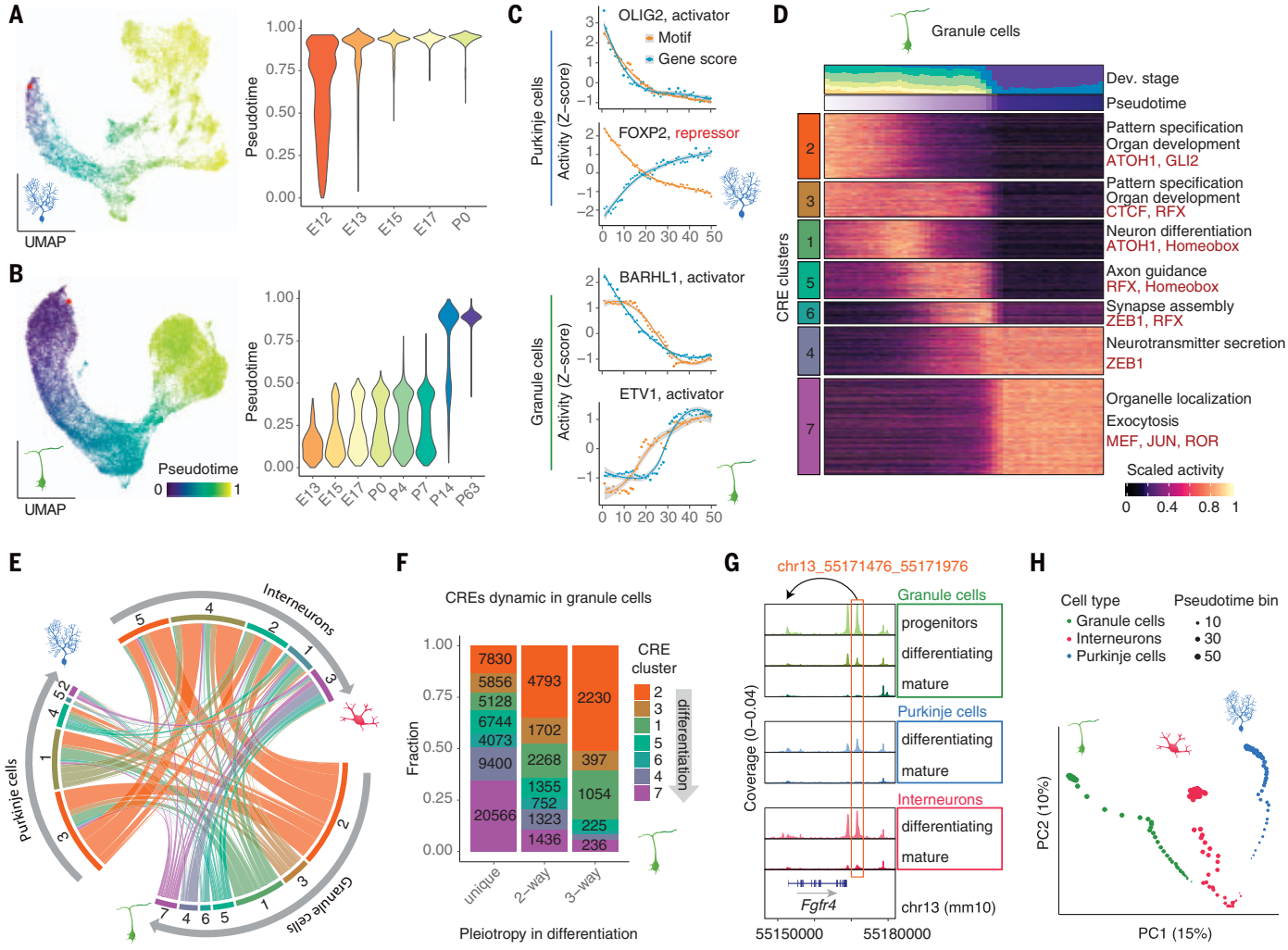


Fig. 4. Gene regulation in neuronal differentiation. (**A** and **B**) Shown on the left are UMAP projections of 13,214 Purkinje cells (A) and 35,153 granule cells (B) aligned across developmental stages, colored by pseudotime value. Red points indicate the pseudotime root. Shown on the right is the distribution of pseudotime values across developmental stages. (**C**) *z*-score scaled values for gene and motif activity across ranked pseudotime bins for examples of activator and repressor TFs in Purkinje cells (top) and granule cells (bottom). Curves are drawn using LOESS regression; gray areas indicate 95% confidence intervals. (**D**) *z*-score scaled activity of dynamic CREs during granule cell differentiation, averaged across 50 bins of increasing pseudotime ranks. The contribution of developmental stages and mean pseudotime value for each bin are shown at the top. Representative enrichments (BH-adjusted $P < 0.05$; hypergeometric test) for biological processes of adjacent genes (black) and TF motifs (red) are shown on the right. (**E**) Overlap between activity clusters for CREs dynamic in two or more neurons (pleiotropic). For each neuron type (outer sector), CRE clusters [as in (D) and fig. S14] are ordered from early (orange) to late (violet) activity during differentiation. Each node connects the activity clusters of two different neuron types for the same CRE. (**F**) Fraction of CRE clusters [as in (D)] across CREs dynamic in a single neuron type (unique) or shared across two or three cell types. (**G**) Example of a pleiotropic intergenic CRE, assigned to *Fgfr4*. Accessibility profiles for each cell type and state (from Fig. 1C) were aggregated across cells from all developmental stages and scaled by the total number of fragments in each group. chr13, chromosome 13. (**H**) Principal components analysis of CRE accessibility during granule cell, Purkinje cell, and interneuron differentiation. Percentage values show the proportion of variance explained by each component.

(ERVs); and L1 elements (fig. S17, A and B). By contrast, astroglial cells (progenitors and astrocytes) show the most conserved intergenic CREs in the mature cerebellum (Fig. 5A and fig. S16A). They are enriched for CREs that originated 160 million to 177 million years ago in common mammalian and therian (i.e., eutherian-marsupial) ancestors (Fig. 5B) and have since been preserved by purifying selection (fig. S17C). We confirmed and extended these observations using a single-cell chromatin accessibility atlas of adult mouse organs (*15*). We found that astrocytes have the most conserved intergenic CREs, not only in the adult cerebellum but also across all adult organs (Fig. 5F). Moreover, 8 of the 10 most conserved cell types across all organs were neural (Fig. 5F), highlighting the overall stronger evolutionary constraints in the brain (*39*, *45*). Consistently, despite having the fastest evolving regulatory landscape in the cerebellum, microglia constitute the most conserved immune cell type (Fig. 5F).

Differences among neurons are subtler than those among glial cells (Fig. 5A and fig. S16A). Granule cells exhibit the lowest conservation levels compared with other stage-matched neuron types, especially during prenatal development (Fig. 5A and fig. S16A; E13: $P < 10^{-15}$, Mann-Whitney *U* test). This difference might reflect the high turnover in developmental adaptations associated with granule cell amplification across vertebrates, including the emergence of the external granule cell layer, a secondary germinal zone specific to amniotes (*3*, *46*). Collectively, our results reveal common temporal trends as well as cell type–specific

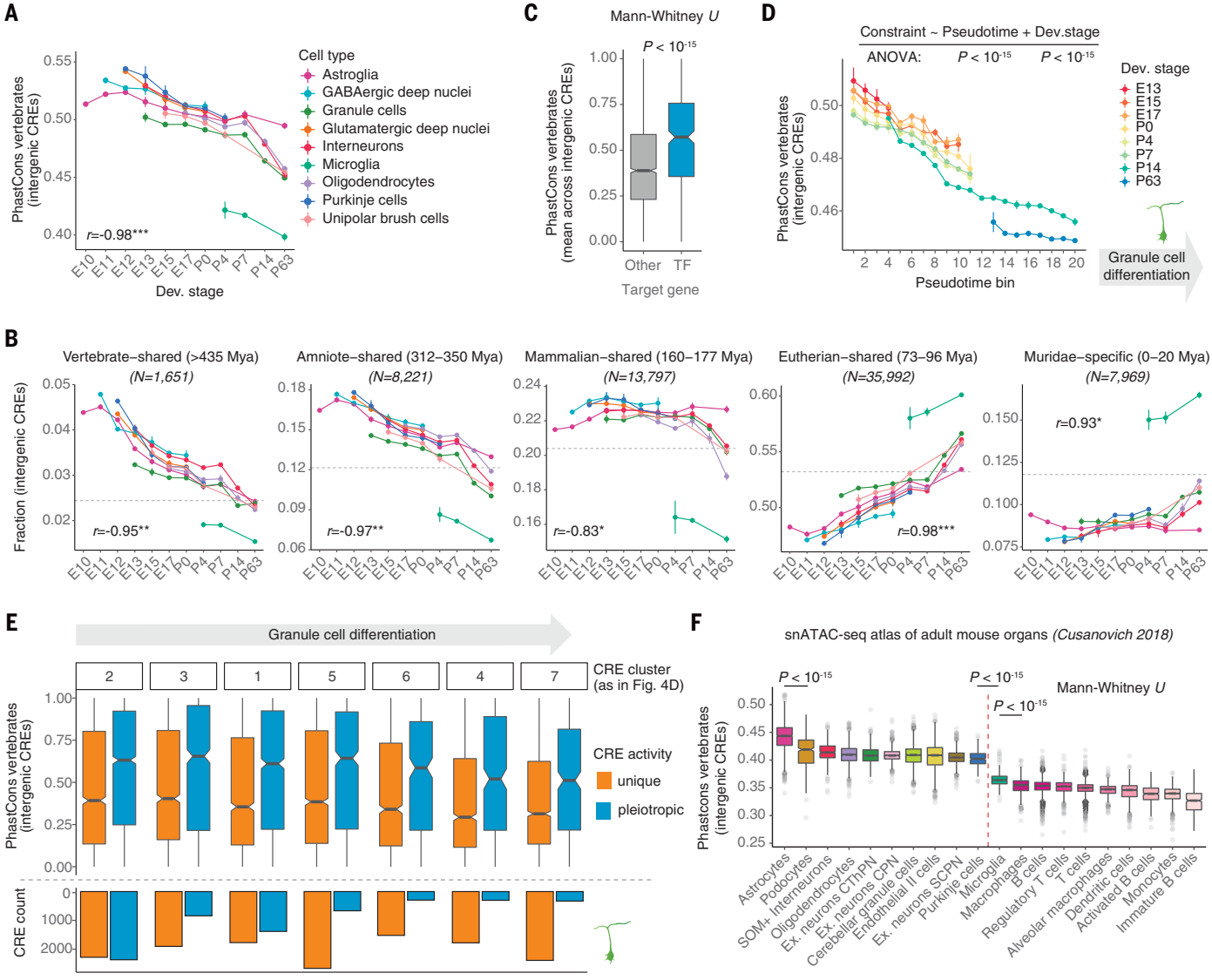


Fig. 5. Evolutionary dynamics of CREs. (**A**) Sequence constraint in intergenic CREs accessible per cell, averaged for each cell type and developmental stage. (**B**) Fraction of accessible intergenic CREs assigned to different age groups per cell, averaged for each cell type and developmental stage. Different *y*-axis ranges were used across age groups to facilitate comparisons between cell types and stages within each group, because the fraction depends on the number of CREs per group (indicated on top). The fraction of each age group across all intergenic CREs is marked by the dotted horizontal line. For (A) and (B), Pearson's *r* correlation coefficients between the estimates and development are shown (median across cell types; **P* < 0.05, ***P* < 0.01, and ****P* < 0.001). Only groups with at least 50 cells were considered. Mya, million years ago. (**C**) Average sequence constraint of intergenic CREs per target gene for TFs and other genes. (**D**) Sequence constraint of intergenic CREs accessible per cell averaged for each developmental stage and pseudotime interval of differentiating granule cells (from Fig. 4B). ANOVA, analysis of variance. (**E**) Sequence constraint (top) and abundance (bottom) of cell type-specific and pleiotropic intergenic CREs active in different stages of granule cell differentiation (from Fig. 4D, ordered from early to late activity during differentiation). (**F**) Sequence constraint of intergenic CREs accessible per cell across cell types in the adult mouse [data from (*15*)]. The 10 most conserved (left) and all immune (right) cell types are shown. SOM+, somatostatin-expressing; Ex. neurons CThPN, excitatory corticothalamic projection neurons; Ex. neurons CPN, excitatory callosal projection neurons; Ex. neurons SCPN, excitatory subcerebral projection neurons. For (A), (B), and (D), error bars represent the difference in average estimates between biological replicates. For (C), (E), and (F), boxes represent the interquartile range, and whiskers extend to extreme values within 1.5 times the interquartile range from the box.

differences in the evolutionary dynamics of CREs in the developing cerebellum.

CRE activity conservation across therian mammals

Although sequence constraint is a strong predictor of CRE activity conservation, comparative studies have shown that regulatory activity, especially for distal elements, evolves faster than DNA sequences (*6*–*8*, *43*). To assess whether our observations on sequence constraints extend to CRE activity conservation, we examined the chromatin accessibility of cerebellar cells in the gray short-tailed opossum, a marsupial separated from mouse by ~160 million years of evolution. We profiled ~20,000 cells spanning two developmental stages: P21, which is transcriptionally similar to P4 in mouse (*39*), and adult (Fig. 6A). We detected the same cell types as in the mouse cerebellum at corresponding stages, with similar abundances,

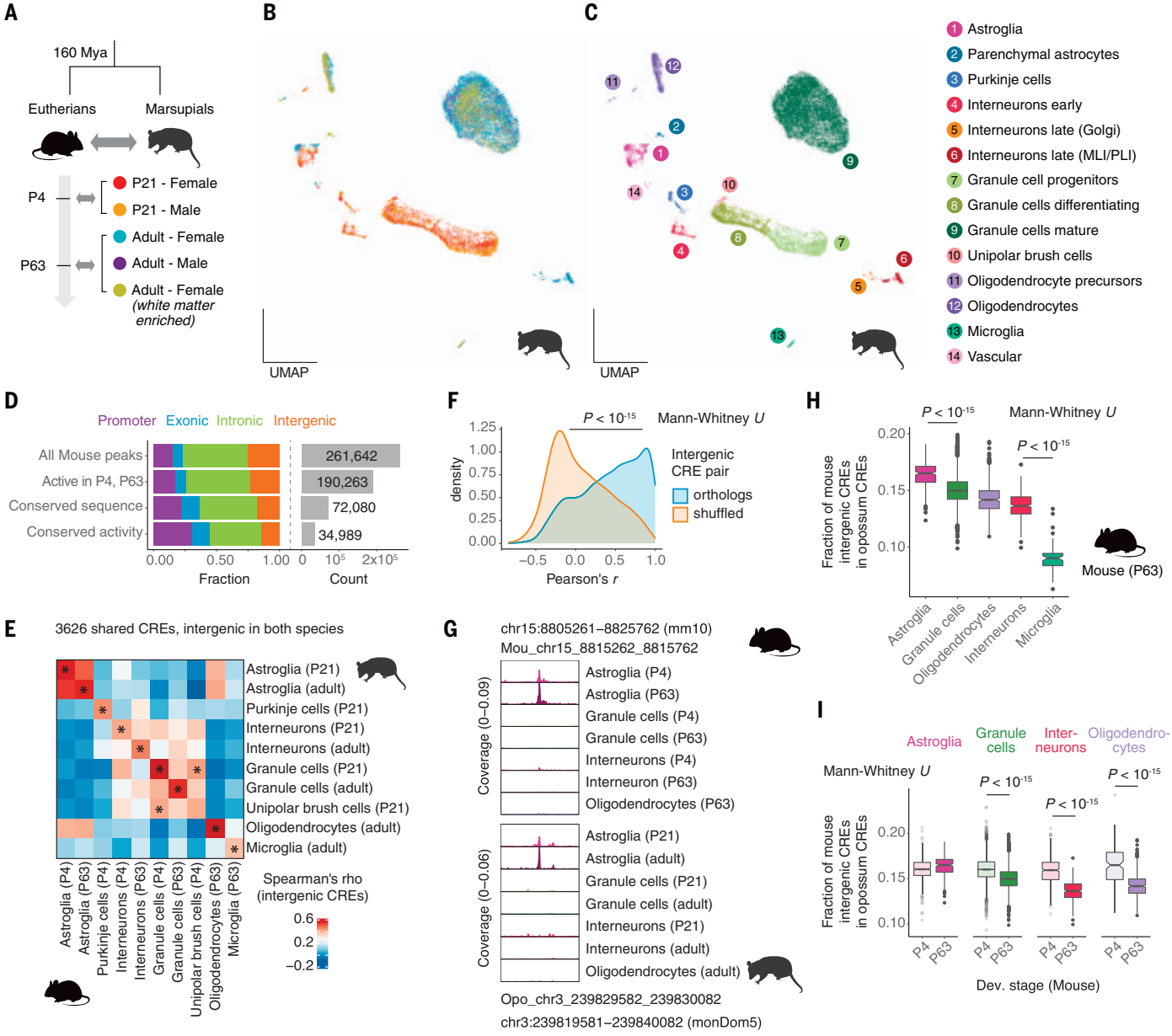


Fig. 6. CRE activity conservation across therian mammals. (**A**) Overview of opossum snATAC-seq dataset and correspondence to mouse developmental stages based on transcriptomic similarity (*39*). (**B** and **C**) UMAP projection of 19,204 opossum cells colored by sample (B) or cell type and state (C). MLI/PLI, molecular/Purkinje cell layer interneurons. (**D**) Distribution of genomic classes (left) and total number (right) for all mouse CREs, CREs active (≥5 CPM) in cell types and developmental stages corresponding to those sampled in opossum, and subsets showing conservation of sequence (CRE aligned to genome) and activity (CRE aligned to CRE) in opossum. (**E**) Spearman's correlation of intergenic CRE activity between opossum and mouse corresponding cell types and stages. Stars mark the sample with the highest correlation for each row and column. (**F**) Pearson's correlation coefficients of activity (in CPM) across corresponding cell types and stages for intergenic CRE pairs with true or shuffled orthology relationships. (**G**) Example of a shared intergenic CRE with conserved cell type–specificity between mouse and opossum. Accessibility profiles for each broad cell type and stage were aggregated across cells and scaled by the total number of fragments in each group. (**H**) Fraction of mouse intergenic CREs accessible per cell with an ortholog CRE in opossum, grouped by cell type in the adult mouse. (**I**) Fraction of mouse intergenic CREs accessible per cell with an ortholog CRE in opossum, grouped by cell type and developmental stage for cell types with at least 100 cells in both stages. For (H) and (I), boxes represent the interquartile range, and whiskers extend to extreme values within 1.5 times the interquartile range from the box.

and with many shared marker genes (Fig. 6, B and C; and fig. S18, A and B). We identified 167,340 putative CREs and compared them with mouse CREs based on reciprocal syntenic alignments. Nearly 50% of mouse CREs with conserved sequences also show conserved activity, with distal CREs evolving faster than promoters (Fig. 6D), as previously described (*6*, *7*). Despite this turnover, all homologous cell types (except for UBCs) show the highest similarity in accessibility profiles of shared intergenic CREs (Fig. 6E), suggesting that most CREs that have been preserved during mammalian evolution maintain similar specificity patterns. Indeed, orthologous intergenic CREs show significantly higher correlations in their activity across matched cell types and developmental stages than shuffled homology pairs

(Fig. 6F and table S7). For example, a mouse CRE assigned to the astroglia marker *Slc1a3* and located 100 kb downstream of the gene has maintained the same astroglia-specific activity in both species (Fig. 6G). Overall, our data suggest that despite the high turnover of regulatory activity in terms of the emergence and loss of CREs (*6*, *7*), radical repurposing of spatiotemporal CRE activity appears to be rare, at least between cell types in the same tissue.

We next reexamined our evolutionary comparisons in consideration of CRE activity conservation. In the adult mouse cerebellum, the fraction of intergenic CREs with a homolog active in the opossum cerebellum is the highest for astrocytes and the lowest for microglia (Fig. 6H), even when requiring the CRE to be active in the corresponding cell type and stage (fig. S18C). Except for astroglia, all cell types show higher conservation in P4 compared with the adult (Fig. 6I), confirming the temporal patterns observed based on sequence conservation (Fig. 5A). Finally, stratifying P4 granule cells according to their differentiation status revealed a significant, albeit small, decline in the fraction of mouse CREs shared with opossum (fig. S18D). Collectively, our analyses of regulatory activity conservation across ~160 million years of mammalian evolution reinforced our sequence-based conclusions regarding differences in CRE constraint across cell types and developmental stages.

Discussion

Our single-cell atlas of gene regulation in the mouse cerebellum spans development from the beginning of neurogenesis to adulthood and is complemented with chromatin accessibility maps for the P21 and adult opossum cerebellum. We provide a comprehensive resource of the cis-regulatory profiles of cerebellar cell types, though not all cell types could be profiled at all stages because of low abundances. We detected extensive temporal differences in CRE activity between cells of the same cell type and even between matched differentiation states. Strong temporal differences in early progenitor cells are shared between germinal zones, suggesting that cell fates are induced through common temporal cues, which could be part of an intrinsic brain-wide temporal code (*33*) or involve common extrinsic signals, such as secreted factors (*3*).

Evolutionary comparisons revealed a decrease in sequence constraint and activity conservation of CREs during development across all cell types. This pattern is explained by both cell-type differentiation and temporal differences between cells from matched differentiation states. The decrease in CRE evolutionary constraint parallels a decline in pleiotropy, as observed in gene expression studies (*39*, *40*). Across cerebellar cell types, differences in regulatory conservation are most pronounced in the adult, where microglia—the immune cells of the brain—show the fastest evolutionary turnover. By contrast, mature astrocytes have the most conserved intergenic CREs, not only in the cerebellum but also across other adult mouse organs. This difference might be explained by astrocytes maintaining some of the properties of neural progenitors, including the ability to reactivate proliferation (*47*), and/or by pleiotropic constraints due to their bridging interactions with multiple cell types of the otherwise disconnected neuronal and vascular networks in the brain (*48*). Alternatively, higher sequence constraints could be associated with a more rigid motif grammar in the regulatory program of astrocytes (*5*). Collectively, our study illuminates how development and evolution shape the cis-regulatory profiles of cerebellar cells and provides general insights into the dynamics of gene regulation during mammalian organ development.

Methods summary

Detailed information on materials and methods can be found in the supplementary materials. In brief, snap-frozen cerebella samples from RjOrl:SWISS mice and gray short-tailed opossums were used for nuclei isolation and single-cell library construction using the Chromium Single Cell ATAC Reagent kits (10x Genomics). Raw sequencing data were processed using cellranger-atac (*21*). ArchR (*22*) was used for quality control, dimensionality reduction, clustering, and identification of open chromatin regions (CREs). Chromosome Y was excluded from all analyses. Chromosome X was excluded from dimensionality reduction but considered for subsequent analyses. Cell-type identification was performed based on ArchR's estimates of gene scores and comparison to known marker genes (*2*, *24*) and in situ hybridization data from the Allen Brain Atlas resource (*49*).

Robust CREs were identified in a cluster-specific manner requiring reproducibility in at least two samples and accessibility in at least 5% of the cells of at least one cluster. Target gene assignment was performed based on promoter-peak coaccessibility and correlations with gene scores. CRE activity across groups (e.g., Fig. 2) was estimated by aggregating accessibility profiles across cells and scaling by sequencing depth (counts per million, CPM). For clustering of CRE activity, CPM values were standardized for each CRE as a fraction of its maximum CPM value across all groups and subjected to k-means clustering, followed by hierarchical clustering of the cluster means. For genes with temporal changes across progenitor types, we estimated average gene scores in groups of cells from the same progenitor type, developmental stage, and replicate. Time-variant genes were identified as those with a higher standard deviation across stages compared with progenitor type and between replicates.

For modeling differentiation trajectories, we used Harmony (*50*) to align across developmental stages and diffusion pseudotime (*51*) to approximate differentiation. For CRE evolutionary dynamics, we estimated sequence constraint as the mean phastCons score (*41*) of the most conserved 100–base pair region for each CRE. CRE minimum age was inferred based on the phylogenetic distance between mouse and the most distant vertebrate species for which a syntenic region could be detected. To estimate average values across cells, we only used intergenic CREs to avoid biases from overlapping or adjacent protein-coding gene sequences. CREs conserved in mouse and opossum were identified based on syntenic alignments between the two species (minimum match of 10%) and requiring reciprocally best matches.

REFERENCES AND NOTES

1. A. Sathyanesan *et al*., Emerging connections between cerebellar development, behaviour and complex brain disorders. *Nat. Rev. Neurosci.* **20**, 298–313 (2019). doi: 10.1038/s41583-019-0152-2; pmid: 30923348
2. K. Leto *et al*., Consensus paper: Cerebellar development. *Cerebellum* **15**, 789–828 (2016). doi: 10.1007/s12311-015-0724-2; pmid: 26439486
3. T. Butts, M. J. Green, R. J. T. Wingate, Development of the cerebellum: Simple steps to make a 'little brain'. *Development* **141**, 4031–4041 (2014). doi: 10.1242/dev.106559; pmid: 25336734
4. A. S. Nord, A. E. West, Neurobiological functions of transcriptional enhancers. *Nat. Neurosci.* **23**, 5–14 (2020). doi: 10.1038/s41593-019-0538-5; pmid: 31740812
5. H. K. Long, S. L. Prescott, J. Wysocka, Ever-changing landscapes: Transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016). doi: 10.1016/j.cell.2016.09.018; pmid: 27863239
6. D. Villar *et al*., Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015). doi: 10.1016/j.cell.2015.01.006; pmid: 25635462
7. C. Berthelot, D. Villar, J. E. Horvath, D. T. Odom, P. Flicek, Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat. Ecol. Evol.* **2**, 152–163 (2018). doi: 10.1038/s41559-017-0377-2; pmid: 29180706
8. S. K. Reilly *et al*., Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015). doi: 10.1126/science.1260943; pmid: 25745175
9. G. Bejerano *et al*., Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004). doi: 10.1126/science.1098119; pmid: 15131266
10. D. E. Dickel *et al*., Ultraconserved enhancers are required for normal development. *Cell* **172**, 491–499.e15 (2018). doi: 10.1016/j.cell.2017.12.017; pmid: 29358049
11. D. U. Gorkin *et al*., An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744–751 (2020). doi: 10.1038/s41586-020-2093-3; pmid: 32728240
12. C. L. Frank *et al*., Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum. *Nat. Neurosci.* **18**, 647–656 (2015). doi: 10.1038/nn.3995; pmid: 25849986
13. A. S. Nord *et al*., Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013). doi: 10.1016/j.cell.2013.11.033; pmid: 24360275
14. A. B. Stergachis *et al*., Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* **154**, 888–903 (2013). doi: 10.1016/j.cell.2013.07.020; pmid: 23953118
15. D. A. Cusanovich *et al*., A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324.e18 (2018). doi: 10.1016/j.cell.2018.06.052; pmid: 30078704

16. Y. E. Li *et al.*, An atlas of gene regulatory elements in adult mouse cerebrum. bioRxiv 2020.05.10.087585 [Preprint] (2020); doi: 10.1101/2020.05.10.087585
17. S. Preissl *et al.*, Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018). doi: 10.1038/s41593-018-0079-3; pmid: 29434377
18. D. J. Di Bella *et al.*, Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* **595**, 554–559 (2021). doi: 10.1038/s41586-021-03670-5; pmid: 34163074
19. R. S. Ziffra *et al.*, Single cell epigenomic atlas of the developing human brain and organoids. bioRxiv 2019.12.30.891549 [Preprint] (2019); doi: 10.1101/2019.12.30.891549
20. S. Domcke *et al.*, A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020). doi: 10.1126/science.aba7612; pmid: 33184180
21. A. T. Satpathy *et al.*, Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019). doi: 10.1038/s41587-019-0206-z; pmid: 31375813
22. J. M. Granja *et al.*, ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021). doi: 10.1038/s41588-021-00790-6; pmid: 33633365
23. M. C. Vladoiu *et al.*, Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* **572**, 67–73 (2019). doi: 10.1038/s41586-019-1158-7; pmid: 31043743
24. J. W. Wizeman, Q. Guo, E. M. Wilion, J. Y. H. Li, Specification of diverse cell types during early neurogenesis of the mouse cerebellum. *eLife* **8**, e42388 (2019). doi: 10.7554/eLife.42388; pmid: 30735127
25. K. J. Millen, E. Y. Steshina, I. Y. Iskusnykh, V. V. Chizhikov, Transformation of the cerebellum into more ventral brainstem fates causes cerebellar agenesis in the absence of Ptf1a function. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1777–E1786 (2014). doi: 10.1073/pnas.1315024111; pmid: 24733890
26. A. Kriegstein, A. Alvarez-Buylla, The glial nature of embryonic and adult neural stem cells. *Annu. Rev. Neurosci.* **32**, 149–184 (2009). doi: 10.1146/annurev.neuro.051508.135600; pmid: 19555289
27. R. A. Carter *et al.*, A single-cell transcriptional atlas of the developing murine cerebellum. *Curr. Biol.* **28**, 2910–2920.e2 (2018). doi: 10.1016/j.cub.2018.07.062; pmid: 30220501
28. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser—A database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007). doi: 10.1093/nar/gkl822; pmid: 17130149
29. M. Dalby, S. Rennie, R. Andersson, FANTOM5 transcribed enhancers in mm10. Zenodo (2018); doi: 10.5281/zenodo.1411211
30. B. Bonev *et al.*, Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017). doi: 10.1016/j.cell.2017.09.043; pmid: 29053968
31. G. Sabarís, I. Laiker, E. Preger-Ben Noon, N. Frankel, Actors with multiple roles: Pleiotropic enhancers and the paradigm of enhancer modularity. *Trends Genet.* **35**, 423–433 (2019). doi: 10.1016/j.tig.2019.03.006; pmid: 31005339
32. W. Meuleman *et al.*, Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251 (2020). doi: 10.1038/s41586-020-2559-3; pmid: 32728217
33. A. Sagner, I. Zhang, T. Watson, J. Lazaro, M. Melchionda, J. Briscoe, Temporal patterning of the central nervous system by a shared transcription factor code. bioRxiv 2020.11.10.376491 [Preprint] (2020); doi: 10.1101/2020.11.10.376491
34. T. Zhang *et al.*, Generation of excitatory and inhibitory neurons from common progenitors via Notch signaling in the cerebellum. *Cell Rep.* **35**, 109208 (2021). doi: 10.1016/j.celrep.2021.109208; pmid: 34107249
35. Y. Seto *et al.*, Temporal identity transition from Purkinje cell progenitors to GABAergic interneuron progenitors in the cerebellum. *Nat. Commun.* **5**, 3337 (2014). doi: 10.1038/ncomms4337; pmid: 24535035
36. V. Cerrato *et al.*, Multiple origins and modularity in the spatiotemporal emergence of cerebellar astrocyte heterogeneity. *PLOS Biol.* **16**, e2005513 (2018). doi: 10.1371/journal.pbio.2005513; pmid: 30260948
37. V. Kozareva *et al.*, A transcriptomic atlas of the mouse cerebellum reveals regional specializations and novel cell types. bioRxiv 2020.03.04.976407 [Preprint] (2020); doi: 10.1101/2020.03.04.976407
38. L. Telley *et al.*, Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science* **364**, eaav2522 (2019). doi: 10.1126/science.aav2522; pmid: 31073041
39. M. Cardoso-Moreira *et al.*, Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019). doi: 10.1038/s41586-019-1338-5; pmid: 31243369
40. I. Sarropoulos, R. Marin, M. Cardoso-Moreira, H. Kaessmann, Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* **571**, 510–514 (2019). doi: 10.1038/s41586-019-1341-x; pmid: 31243368
41. A. Siepel *et al.*, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005). doi: 10.1101/gr.3715005; pmid: 16024819
42. D. Arendt *et al.*, The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016). doi: 10.1038/nrg.2016.127; pmid: 27818507
43. M. Roller *et al.*, LINE retrotransposons characterize mammalian tissue-specific and evolutionarily dynamic regulatory regions. *Genome Biol.* **22**, 62 (2021). doi: 10.1186/s13059-021-02260-y; pmid: 33602314
44. L. Geirsdottir *et al.*, Cross-species single-cell analysis reveals divergence of the primate microglia program. *Cell* **179**, 1609–1622.e16 (2019). doi: 10.1016/j.cell.2019.11.010 pmid: 31835035
45. D. Brawand *et al.*, The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011). doi: 10.1038/nature10532; pmid: 22012392
46. A. Iulianella, R. J. Wingate, C. B. Moens, E. Capaldo, The generation of granule cells during the development and evolution of the cerebellum. *Dev. Dyn.* **248**, 506–513 (2019). doi: 10.1002/dvdy.64; pmid: 31131952
47. M. Götz, S. Sirko, J. Beckers, M. Irmler, Reactive astrocytes as neural stem or progenitor cells: In vivo lineage, In vitro potential, and Genome-wide expression analysis. *Glia* **63**, 1452–1468 (2015). doi: 10.1002/glia.22850; pmid: 25965557
48. A. Volterra, J. Meldolesi, Astrocytes, from brain glue to communication elements: The revolution continues. *Nat. Rev. Neurosci.* **6**, 626–640 (2005). doi: 10.1038/nrn1722; pmid: 16025096
49. Allen Institute for Brain Science, Allen Developing Mouse Brain Atlas (2008); http://developingmouse.brain-map.org/.
50. I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019). doi: 10.1038/s41592-019-0619-0; pmid: 31740819
51. L. Haghverdi, M. Büttner, F. A. Wolf, F. Buettner, F. J. Theis, Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016). doi: 10.1038/nmeth.3971; pmid: 27571553

ACKNOWLEDGMENTS

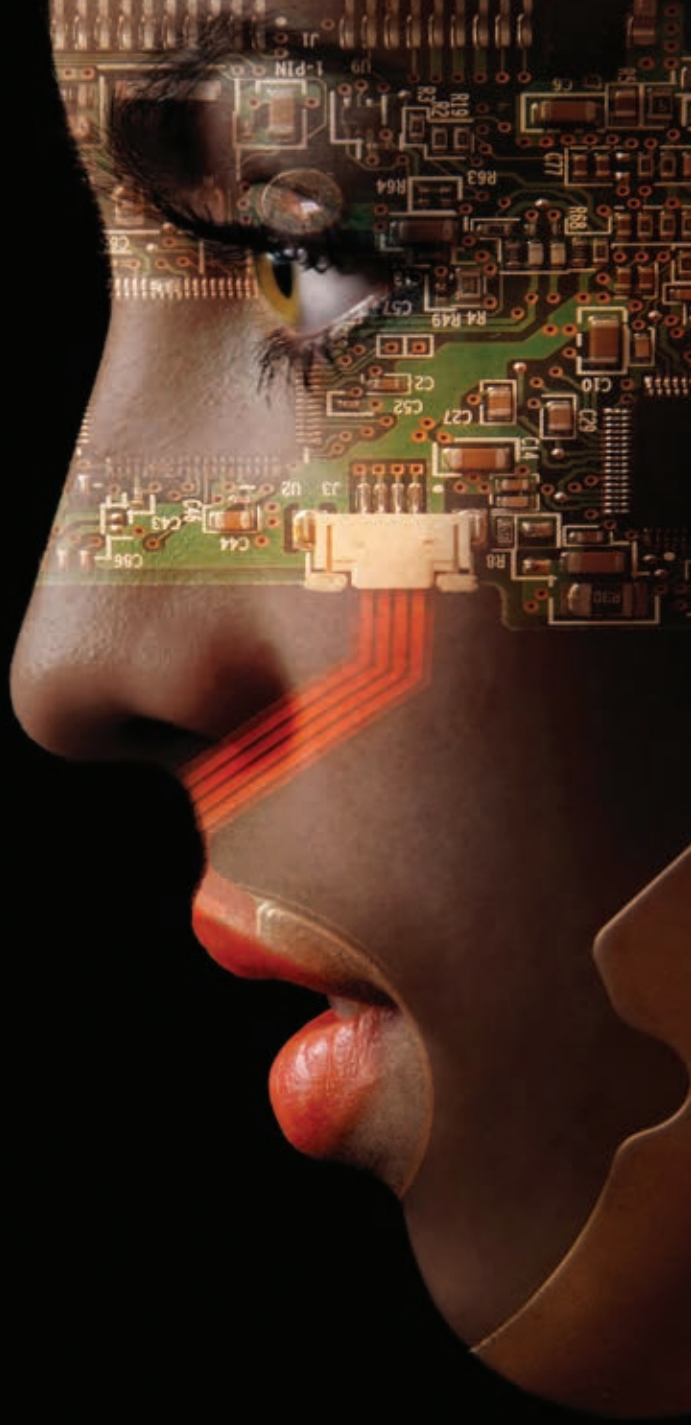
We thank D. Odom and S. M. Waszak for critical reading of the manuscript and S. Anders, D. Kawauchi, F. Lamanna, F. Murat, O. Stegle, J. Zaugg, and members of the Kaessmann lab for discussions. We thank J. Schmidt, C. Schneider, A. Billepp, and P. Grimm for technical support. **Funding:** This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 819894) and Seventh Framework Programme (FP7-2007-2013) (grant agreement no. 615253). M.C.-M. was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC011171), the UK Medical Research Council (FC011171), and the Wellcome Trust (FC011171). **Author contributions:** H.K., M.C.-M., I.S., and M.S. originally designed the study; M.S. collected samples; P.G. maintained a breeding colony of opossums; M.S. and R.F. generated data; I.S. performed all analyses with contributions from M.S. and key input from R.F., K.L., and N.T.; E.L., L.M.K., K.O., and P.J. provided important feedback and discussions; H.K. and S.M.P. provided funding; H.K., S.M.P., and M.C.-M. supervised the study; and I.S., M.S., M.C.-M., and H.K. wrote the manuscript with input from all authors. **Competing interests:** The authors report no conflicts of interest. **Data and materials availability:** All data generated in this study are freely accessible in ArrayExpress with the accession codes E-MTAB-9765 (mouse) and E-MTAB-10533 (opossum). All other data are in the main paper or the supplementary materials. All code used to analyze the data is available online at https://github.com/ioansarr/mouse-cerebellum-atac/. Processed data can be interactively explored at https://apps.kaessmannlab.org/mouse_cereb_atac/. Active CREs across cell types and stages (methods) can be inspected as custom UCSC genome browser tracks at: http://genome.ucsc.edu/s/ioansarr/Mouse%20cerebellum%20snATAC-seq%20cCREs. This research was funded in part by Wellcome Trust (FC011171), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/eabg4696/suppl/DC1
Materials and Methods
Figs. S1 to S18
Tables S1 to S7
References (*52*–*93*)
MDAR Reproducibility Checklist

8 January 2021; accepted 14 July 2021
Published online 29 July 2021
10.1126/science.abg4696

ScienceRobotics.org



DOESN'T YOUR RESEARCH DESERVE THE BEST READERS?

Submit your research: **cts.ScienceMag.org**

Science Robotics
AAAS

Twitter: @SciRobotics
Facebook: @ScienceRobotics

RESEARCH ARTICLES

MOLECULAR BIOLOGY

RNA editing restricts hyperactive ciliary kinases

Dongdong Li[1,2,3,4], Yufan Liu[1,2,3,4], Peishan Yi[1,2,3,4], Zhiwen Zhu[1,2,3,4], Wei Li[5], Qiangfeng Cliff Zhang[1,2,6,7], Jin Billy Li[8], Guangshuo Ou[1,2,3,4]*

Protein kinase activity must be precisely regulated, but how a cell governs hyperactive kinases remains unclear. In this study, we generated a constitutively active mitogen-activated protein kinase DYF-5 (DYF-5CA) in *Caenorhabditis elegans* that disrupted sensory cilia. Genetic suppressor screens identified that mutations of ADR-2, an RNA adenosine deaminase, rescued ciliary phenotypes of *dyf-5CA*. We found that *dyf-5CA* animals abnormally transcribed antisense RNAs that pair with *dyf-5CA* messenger RNA (mRNA) to form double-stranded RNA, recruiting ADR-2 to edit the region ectopically. RNA editing impaired *dyf-5CA* mRNA splicing, and the resultant intron retentions blocked DYF-5CA protein translation and activated nonsense-mediated *dyf-5CA* mRNA decay. The kinase RNA editing requires kinase hyperactivity. The similar RNA editing–dependent feedback regulation restricted the other ciliary kinases NEKL-4/NEK10 and DYF-18/CCRK, which suggests a widespread mechanism that underlies kinase regulation.

Protein kinases mediate signal transduction and regulate fundamental cellular processes in eukaryotic cells. The catalytic activity of kinases must be stringently controlled, and its dysregulation leads to developmental defects, metabolic disorders, and cancer (*1*, *2*). Although loss-of-function kinase mutations have been extensively studied, gain-of-function mutations are less understood. Kinase hyperactivities have been described in many disease conditions, such as the constitutively active BCR-ABL1 tyrosine kinase in chronic myelogenous leukemia and BRAF serine/threonine kinase in malignant melanomas (*1*). Potent kinase inhibitors treat cancer effectively; however, tumors inevitably develop drug resistance and circumvent kinase inhibition by activating alternative pathways that perpetuate cell proliferation (*3*). Therefore, it is crucial to understand how an organism regulates hyperactive kinases.

We studied male germ cell–associated kinase (Mak), an evolutionarily conserved mitogen-activated protein kinase (MAPK). Mak localizes at the ciliary tip to restrict cilia elongation (*2*, *4*, *5*). Cilia are microtubule-based organelles essential for cell motility and sensory signaling, and ciliary defects cause more than 35 ciliopathies (*6*). The loss of Mak orthologs results in abnormally long cilia and leads to retinitis pigmentosa (*2*, *4*, *5*, *7*, *8*). By contrast, overexpression of mammalian Mak or *Caenorhabditis elegans* genomic DNA (gDNA) encoding the Mak ortholog DYF-5 inhibits ciliogenesis (*2*, *5*), which indicates that an optimal Mak/DYF-5 activity is essential. The MAPK superfamily contains a canonical TxY motif in the activation T-loop, phosphorylation of the threonine (pThr) residue activates the kinase, and glutamate can be a phosphomimetic pThr (*9*, *10*).

[1]Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing, China. [2]Beijing Frontier Research Center for Biological Structure, Tsinghua University, Beijing, China. [3]McGovern Institute for Brain Research, Tsinghua University, Beijing, China. [4]School of Life Sciences and MOE Key Laboratory for Protein Science, Tsinghua University, Beijing, China. [5]School of Medicine, Tsinghua University, Beijing, China. [6]MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing, China. [7]Center for Synthetic and Systems Biology, Tsinghua University, Beijing, China. [8]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA.
*Corresponding author. Email: guangshuoou@tsinghua.edu.cn

Results

A constitutively active DYF-5 kinase disrupted cilia

We first established that substitution of Thr^{164} with glutamate (T164E) in the TxY motif of DYF-5 leads to constitutive kinase activation in vitro. DYF-5(T164E) phosphorylated an intraflagellar transport (IFT) protein IFT-74, whereas wild-type (WT) DYF-5 did not (fig. S1, A to D), demonstrating that the substitution produced a constitutively active DYF-5 kinase (hereafter DYF-5CA). We then generated a T164E *dyf-5CA* knock-in animal in *C. elegans* (Fig. 1A and fig. S2, A and B). The *dyf-5CA* mutant (n > 200) exhibited normal morphology, viability, and growth rates comparable with those of the WT N2 strain. However, 87% of *dyf-5CA* worms (n = 150) could not use their sensory cilia to uptake fluorescence dye 1,1′-dioctadecyl-3,3,3′,3′-tetramethylindocarbocyanine perchlorate from the medium (Dyf) (Fig. 1B and fig. S2C), indicating an impaired ciliary function and integrity (*11*). To visualize IFT and cilia, we introduced a green fluorescence protein (GFP)–tagged IFT52/OSM-6 into *dyf-5CA*. In contrast to the bidirectional motility of OSM-6::GFP in WT cilia (*12*), no IFT was detectable from the distal ciliary segments of *dyf-5CA* (n > 20 animals), in which the OSM-6::GFP fluorescence formed aggregates (Fig. 1, C and D, and fig. S3, A and C to E).

The *dyf-5CA* animals developed abnormally elongated cilia, resembling the long cilia defect in the *dyf-5(mn400)* null allele (Fig. 1, B and D, and fig. S3B) but opposite to the truncated cilia phenotype when *dyf-5* gDNA was overexpressed (*2*). None of the *dyf-5CA* heterozygotes showed ciliary defects (n > 50 examined heterozygotes), and expression of the *dyf-5* cDNA in ciliated neurons fully rescued ciliary phenotypes in *dyf-5CA* (Fig. 1, B and D), indicating that *dyf-5CA* is a recessive loss-of-function mutant. Expression of *dyf-5CA* cDNA did not disrupt cilia in WT but rescued ciliary defects in *dyf-5CA* mutants (Fig. 1B and fig. S2D). Thus, the DYF-5CA protein did not have a dominant-negative effect. The discrepancy between the in vitro kinase hyperactivity and the in vivo kinase-null phenotype suggests that a previously unrecognized mechanism may control DYF-5CA in a living organism.

RNA adenosine deaminase mutations rescued ciliary defects by DYF-5CA kinase

We performed a forward genetic suppressor screen for mutations that restore *dyf-5CA* cilia (fig. S4A) and isolated four suppressors, *cas506*, *cas517*, *cas518*, and *cas519*, which exhibited uptake of fluorescence dye like the WT (n = 100) (fig. S4, B and C). All four mutations were mapped to the deaminase domain of ADR-2, with *cas517* carrying a premature stop codon and the other three mutants carrying missense mutations (Fig. 1E and fig. S4, B and D). A deletion mutant of *adr-2(ok735)* that removes the entire double-stranded RNA (dsRNA) binding domain and most of the deaminase domain (Fig. 1E and fig. S4E) showed the same ciliary rescue effects on *dyf-5CA* mutant cilia (Fig. 1, B and D, and fig. S3B).

ADR-2 is an ortholog of human ADARB1 (adenosine deaminase acting on RNA B1, or ADAR2). It exhibits dsRNA adenosine deaminase activity, catalyzing the adenosine-to-inosine (A-to-I) RNA editing, which is interpreted by the ribosome as guanosine (*13*, *14*). The *C. elegans* genome encodes two ADARs: ADR-1 is catalytically inactive, and ADR-2 is the only active RNA deaminase. The *adr-2* deletion completely abolishes A-to-I editing (*15*–*17*). Although ADR-1 regulates the activity of ADR-2 (*15*), the *adr-1* deletion did not recover *dyf-5CA* cilia (fig. S4C). By contrast, the *adr-2* deletion reduced the ciliary aggregates of OSM-6::GFP and improved IFT in both amphid and phasmid cilia of *dyf-5CA* animals, albeit at a slower speed (Fig. 1D and fig. S3, A and C to E). None of the *adr-2* mutant alleles displayed ciliary defects (n > 100, for each genotype) (Fig. 1B and fig. S4C). We used a ciliated neuron-specific promoter, P*dyf-1*, to express *adr-2* cDNA in the

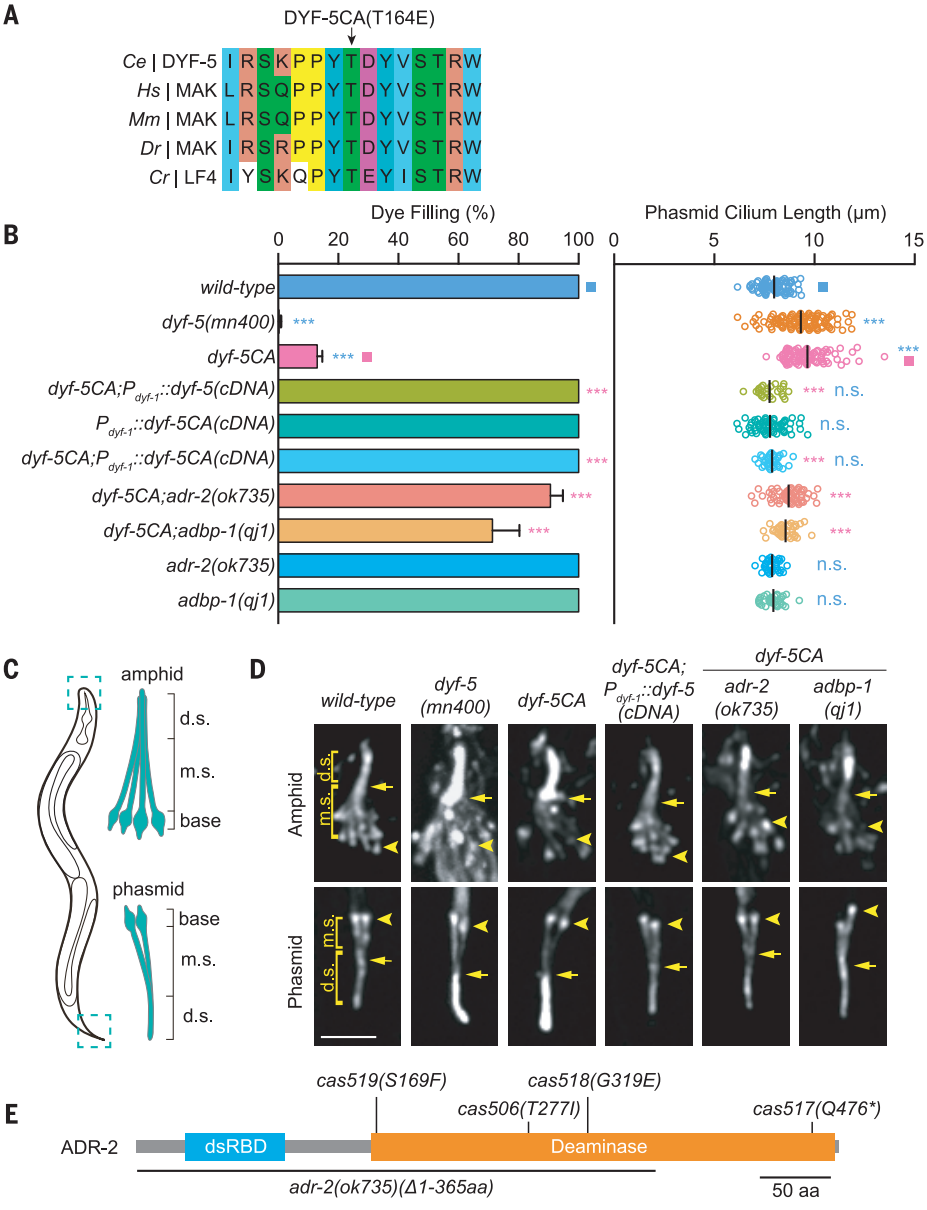


Fig. 1. ADR-2 mutations rescued ciliary defects in *dyf-5CA*. (**A**) Multi-sequence alignment of the *C. elegans* (Ce) DYF-5 kinase and its homologs in human (Hs), mouse (Mm), zebrafish (Dr), and *Chlamydomonas* (Cr). E, Glu; T, Thr. (**B**) Percentage of dye-filling positive animals in the indicated strains (mean ± SD) (left). N = 100 to 200. The cilium length of phasmid cilia (right, measured using the OSM-6::GFP fluorescence) is shown. N = 30 to 60. Statistical significance, compared with the control with a matching color code, is based on Student's *t* test; n.s., not significant, ***$P < 0.001$. (**C**) Schematic of the amphid and phasmid cilia. Dashed boxes are enlarged on the right. d.s., distal ciliary segment; m.s., middle segment. (**D**) Cilia in WT and mutants. Arrowheads indicate the ciliary base and transition zone, and arrows indicate junctions between the middle and distal segments. Scale bar, 5 μm. (**E**) ADR-2 protein domains. Amino acid changes and the deletion in *adr-2* mutant alleles are indicated. dsRBD, dsRNA binding domain. F, Phe; G, Gly; I, Ile; Q, Gln; S, Ser. Scale bar, 50 amino acids (aa).

dyf-5CA; adr-2 double mutant, which caused 75% of animals to develop the same ciliary defects that were observed in *dyf-5CA* animals (n = 300 transgenic animals) (fig. S4C), indicating that ADR-2 functions cell autonomously.

DYF-5CA caused aberrant RNA editing of the dyf-5CA *pre-mRNA*

We sought a mechanistic understanding between ADR-2 and DYF-5CA. An ADR-2::GFP reporter confirmed its nuclear localization (*13*) but did not reveal the ciliary localization of ADR-2 (fig. S5, A and B). Using GFP-affinity purification and mass spectrometry, we identified a known ADR-2 binding protein, ADBP-1, required for its deaminase activity (fig. S5C and data file S1) (*18*). The *adbp-1(qj1)* mutant recovered *dyf-5CA*'s ciliary defects (Fig. 1, B and D, and fig. S3, A to C), and GFP::ADBP-1 localized within nuclei, but not cilia (fig. S5D), suggesting that ADR-2 and ADBP-1 do not directly function in cilia but regulate DYF-5CA by their nuclear actions. Because ADR-2 and ADBP-1 are involved in small interfering RNA– and microRNA-associated pathways (*18*–*20*), we introduced RNA interference (RNAi) mutants into *dyf-5CA* but did not observe rescue effects (fig. S5E), indicating that the function of *adr-2* or *adbp-1* is independent of the RNAi machinery.

To understand ADR-2–based RNA editing in *dyf-5CA*, we implemented a comparative RNA-sequencing (RNA-seq) screen on WT, *dyf-5CA*, *adr-2*, and *dyf-5CA; adr-2* animals by using an established computational pipeline (*21*). WT animals had a total of 18,097 A-to-I editing sites, but RNA editing events were almost undetectable in *adr-2* or *adbp-1* (Fig. 2A, fig. S6A, and data file S2), which confirmed previous findings (*15*–*18*). In *dyf-5CA* animals, 12,871 sites that were edited in WT became unedited (Fig. 2A and data file S2). Because removing all the editing sites in the *dyf-5CA; adr-2* double mutant rescued *dyf-5CA*'s ciliary defects, we argue that these editing sites lost in *dyf-5CA* are not responsible for the phenotypes, although it is unclear why and how these sites were lost. The *dyf-5CA* animals unexpectedly acquired 1537 RNA editing sites, which were not detected in the WT and were lost in the *dyf-5CA; adr-2* (Fig. 2A and data file S2). Most of the identified RNA editing sites are located on chromosomes as clusters (*17*). We searched for the groups of ectopic editing sites and found that clustered sites were present in seven genes (fig. S6B and data file S2), including the *dyf-5CA* gene itself (Fig. 2B). The 269 editing

sites were distributed from the sixth exon to the seventh intron of *dyf-5CA* and occurred at frequencies of 10 to ~60% (Fig. 2C). By contrast, we did not detect any RNA editing sites in the *dyf-5* locus in other strains (Fig. 2B). Except for *dyf-5*, none of the other six genes were known to have ciliary functions.

We next investigated whether and to what extent the aberrant RNA editing clusters in *dyf-5CA* affect cilia. The sixth and seventh introns contain 130 editing sites, and the deletion of either intron or both in *dyf-5CA* animals rescued ciliary defects (Fig. 2D and figs. S2A and S7, A and B). The double intron deletion removed all the ectopic editing clusters in *dyf-5* exons (Fig. 2B). To rule out the effects of intron deletion on pre-mRNA, we synthesized the sixth and seventh introns by placing a noneditable T at all detected A-to-I editing sites. Replacing introns with noneditable sites rescued the ciliary phenotypes in *dyf-5CA* (Fig. 2D and figs. S2A and S7, B and C). Although other aberrantly gained or lost RNA editing sites may have biological functions under certain conditions, our data show that RNA editing sites in *dyf-5CA* are the primary cause for ciliary defects.

dyf-5CA *animals generated antisense transcripts in* dyf-5

Why does RNA editing occur specifically in this region? Because substrates of RNA editing enzymes are dsRNA (*13*), we examined whether long dsRNAs are formed intramolecularly at this locus but failed to identify any, which is consistent with lacking RNA editing at this

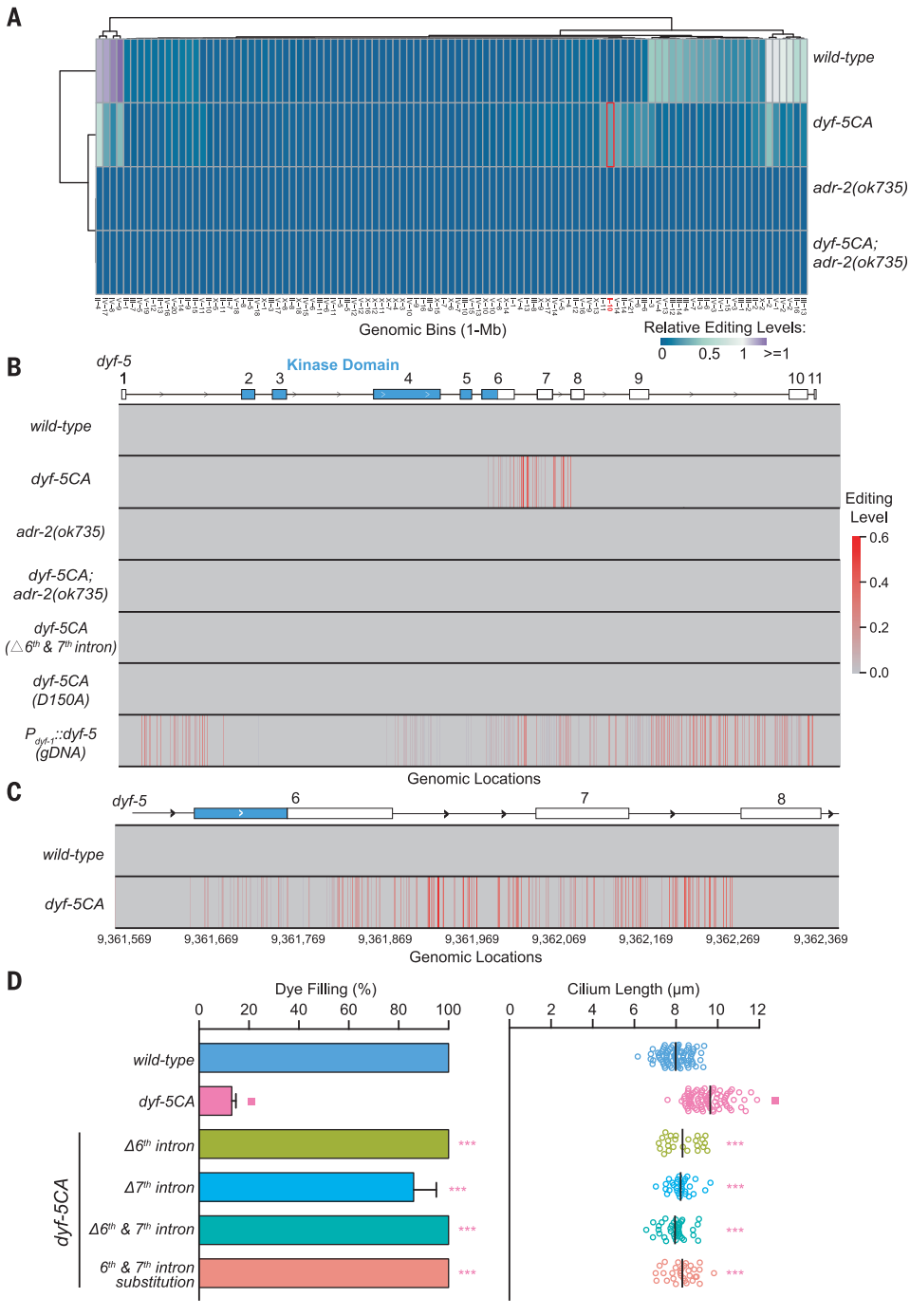


Fig. 2. ADR-2–dependent RNA editing clusters in *dyf-5CA*. (**A**) Heatmap and dendrogram of hierarchical clustering of editing sites using editing ratios among WT and mutants in different genomic bins (1-Mb tile) from RNA-seq data. Relative editing levels were used in hierarchical clustering. Each row corresponds to a sample. Each column represents the relative editing ratio in a genomic bin. Experiments were repeated three times with the similar results. The red box highlights the bin with the abnormally gained RNA editing sites of the *dyf-5* locus in *dyf-5CA*. (**B**) Representative RNA editing analysis at the *dyf-5* locus. Each row corresponds to a sample. Red lines represent RNA editing sites in the genomic locations of *dyf-5*, and the bar for editing level is shown on the right. Gene bodies are above; blue boxes show exons encoding the kinase domain. A, Ala; D, Asp. (**C**) Ectopic gain of RNA editing sites at the *dyf-5* locus in *dyf-5CA*. (**D**) Percentage of dye-filling (mean ± SD) positive animals in the indicated strains, *n* = 100 to 200 (left). Cilium length, *n* = 30 to 60 (right), is shown.

locus in the WT. We then investigated whether an antisense RNA was generated to form dsRNA with the *dyf-5CA* transcript subject to ADR-2 editing. Our strand-specific RNA-seq detected abundant antisense transcripts in *dyf-5CA* animals, starting from the seventh intron to the fourth exon of *dyf-5CA* and aligning with the region where aberrant RNA editing occurred (Fig. 3, A and B). Next, we examined whether the eighth intron of *dyf-5*, which is immediately upstream of the antisense transcripts, regulates antisense transcription. Although *dyf-5* cDNA expression did not cause ciliary abnormalities (Fig. 3C and figs. S2D and S8A), expression of *dyf-5* cDNA containing the sixth, seventh, and eighth introns, but not any individual intron, disrupted cilia (Fig. 3C). Conversely, *dyf-5* gDNA expression impaired cilia in an ADR-2–dependent manner (Fig. 3C and fig. S8A); however, expression of *dyf-5* gDNA that lacked the sixth, seventh, or eighth intron reduced the penetrance of ciliary defects (Fig. 3C). Consistently, *dyf-5* gDNA expression produced ectopic RNA editing clusters and antisense RNA transcription of *dyf-5*, which were not limited to the regions in *dyf-5CA* animals but spread to additional introns and exons of *dyf-5* (Figs. 2B and 3A), generating more profound effects than *dyf-5CA*. Deleting the eighth intron reduced the ciliary defects of *dyf-5CA* animals (Fig. 3C and fig. S7B), which demonstrates its contribution to ciliary defects. The incomplete penetrance of the eighth intron deletion suggests that other genomic region(s) might be involved. This intron may serve as a cryptic promoter for antisense transcription, and future studies will need to identify the additional cis-regulatory elements and the corresponding trans factor(s).

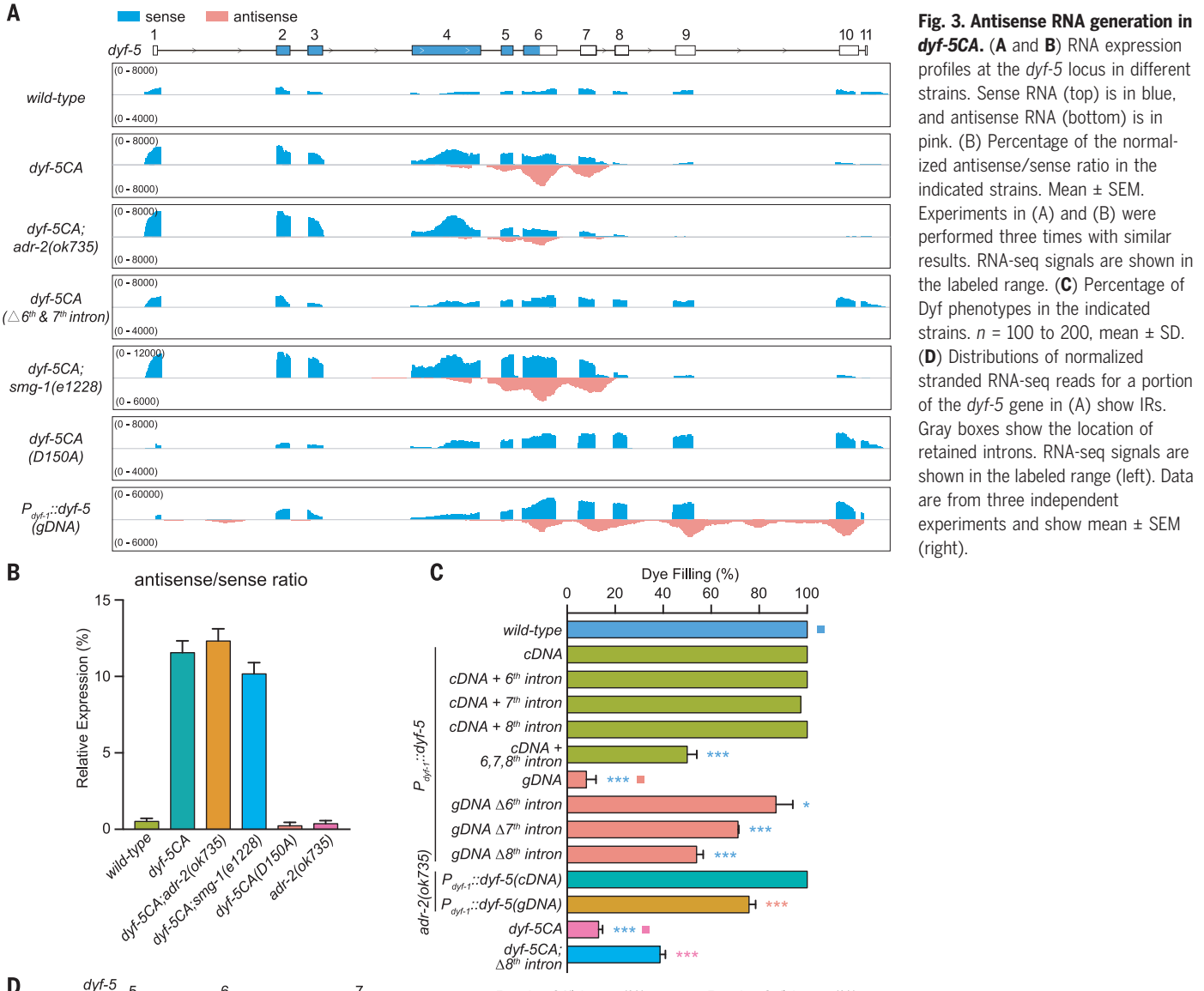


Fig. 3. Antisense RNA generation in *dyf-5CA*. (**A** and **B**) RNA expression profiles at the *dyf-5* locus in different strains. Sense RNA (top) is in blue, and antisense RNA (bottom) is in pink. (B) Percentage of the normalized antisense/sense ratio in the indicated strains. Mean ± SEM. Experiments in (A) and (B) were performed three times with similar results. RNA-seq signals are shown in the labeled range. (**C**) Percentage of Dyf phenotypes in the indicated strains. *n* = 100 to 200, mean ± SD. (**D**) Distributions of normalized stranded RNA-seq reads for a portion of the *dyf-5* gene in (A) show IRs. Gray boxes show the location of retained introns. RNA-seq signals are shown in the labeled range (left). Data are from three independent experiments and show mean ± SEM (right).

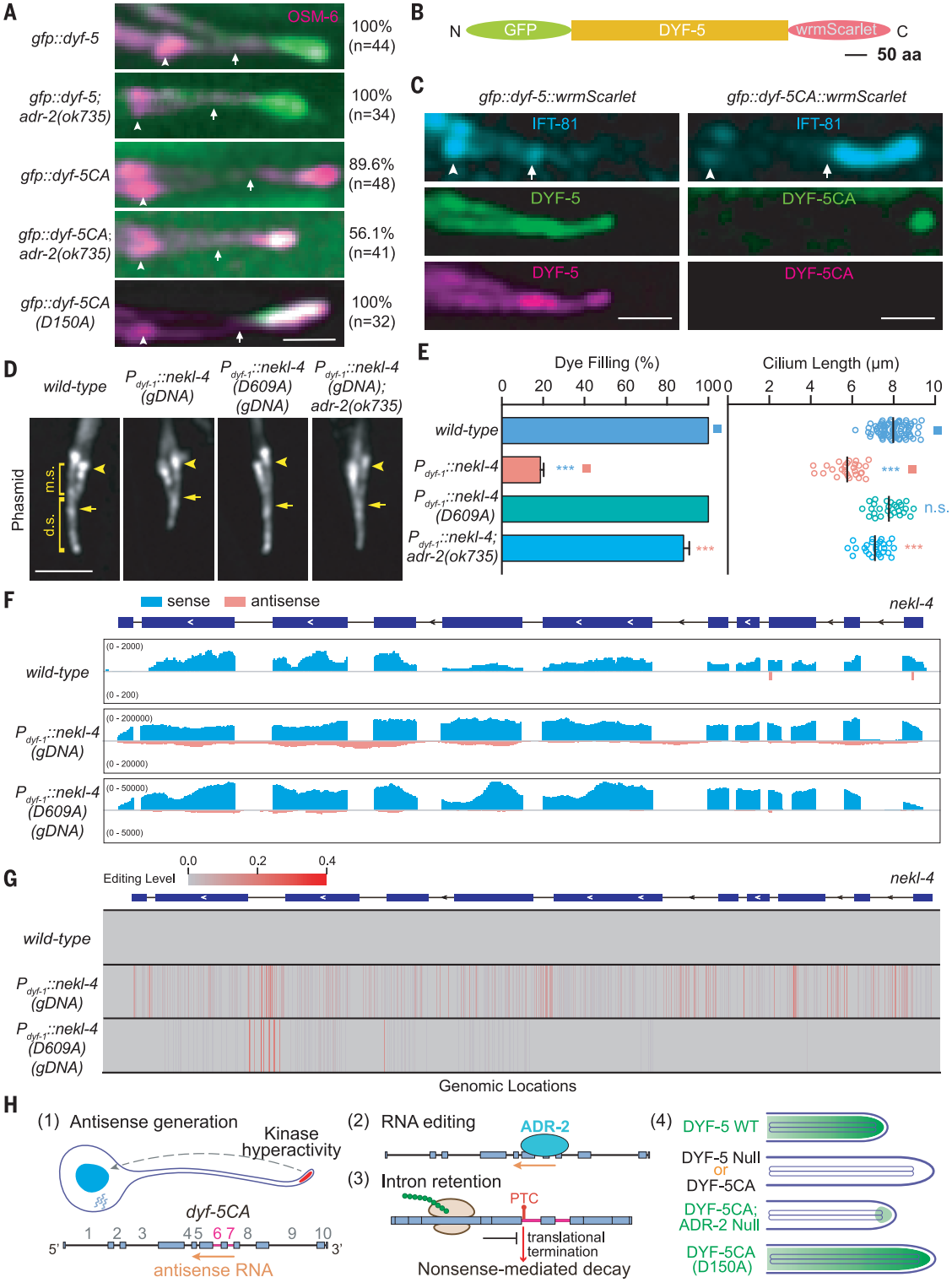


Fig. 4. DYF-5CA protein production and a working model. (**A**) Localization of GFP::DYF-5 or GFP::DYF-5CA with different genotypes. OSM-6::mCherry marks cilia. Percentages of DYF-5 localization patterns are indicated. (**B**) Schematic of *gfp::dyf-5::wrmScarlet* knock-in allele. Scale bar, 50 amino acids. (**C**) Localization of GFP::DYF-5::wrmScarlet (left) or GFP::DYF-5CA::wrmScarlet (right) in phasmid cilia. The blue fluorescence protein (BFP)–tagged IFT-81 labels cilia. Scale bars, (A) and (C), 2.5 μm. (**D**) Ciliary defects in the *nekl-4* gDNA overexpressed animals. Scale bar, 5 μm. (**E**) Percentage of dye filling (mean ± SD)–positive animals, *n* = 100 to 200 (left). Cilium length, *n* = 20 to 40 (right). (**F**) Mapping profiles of sense and antisense RNA at *nekl-4* in the WT *nekl-4* (upper) or kinase-dead *nekl-4(D609A)* (lower) gDNA overexpression animals (sense reads, blue; antisense reads, pink). (**G**) RNA editing clusters at *nekl-4*. The bar for editing level is shown on the top. (**H**) Model summarizes how RNA editing restricts hyperactive ciliary kinases.

Impaired dyf-5CA *mRNA splicing blocked DYF-5CA kinase translation*

Whereas antisense transcripts resulted in promiscuous RNA editing in *dyf-5CA*, it is the RNA editing, but not the antisense transcripts, that caused ciliary defects because *dyf-5CA; adr-2* double mutants that retained antisense transcripts but lacked RNA editing restored their cilia (Fig. 3, A and B). Thus, we studied the molecular consequence that RNA editing conferred in *dyf-5CA*. The hyperediting of the sixth and the seventh exons changed 40.5% of the amino acids in this region. The sixth exon is located in the kinase domain, implying a possible down-regulation of kinase activity (fig. S7D). The sixth and the seventh introns in the *dyf-5CA* pre-mRNA were not sufficiently spliced out, causing intron retention (IR) (Fig. 3, A and D). Using IRFinder (*22*), we found that the sixth and seventh IR ratios are 5.7 and 34.6% in *dyf-5CA*, respectively (Fig. 3D). By contrast, IR ratios were reduced to 0.6 and 16% in the *dyf-5CA;adr-2* double mutant, indicating that RNA editing of the intron sequence may promote IR. The incomplete elimination of IR in *dyf-5CA;adr-2* double mutants suggests that other mechanisms may be involved. For example, RNA editing is known to occur cotranscriptionally in nascent RNAs (*23*), and the ectopic antisense RNA transcripts can form dsRNA to prevent the spliceosome from accessing the *dyf-5CA* pre-mRNA.

To examine kinase production, we fused GFP to DYF-5 at its N terminus in WT or *dyf-5CA* animals. DYF-5 localized in the WT or *adr-2* mutant cilia (100%, $n > 30$ animals) (Fig. 4A and fig. S8B); however, GFP fluorescence was barely detected from 89.6% of *gfp::dyf-5CA* animals (Fig. 4A and fig. S8C). In *gfp::dyf-5CA;adr-2* double mutants, the GFP::DYF-5CA signal increased, albeit at a lower level than the WT (Fig. 4A), indicating that a small amount of hyperactive kinase restored cilia. To further evaluate whether the loss of GFP::DYF-5CA resulted from translation blockade, we fused a red fluorescent protein gene (wrmScarlet, the Scarlet protein with worm codon optimization) to the C terminus of GFP::DYF-5 or GFP::DYF-5CA (Fig. 4B). The green and red fluorescence colocalized in all the *gfp::dyf-5::wrmScarlet* cilia ($n = 60$ animals) (Fig. 4C); however, only the faint green but no red fluorescence was visible from all the examined *gfp::dyf-5CA::wrmScarlet* animals ($n = 43$) (Fig. 4C), which implies premature translation termination. Furthermore, the GFP-tagged DYF-5 kinase domain (11 to 291 amino acids) was localized in the cell body and dendrite but not in cilia, whereas the C terminus of DYF-5 (292 to 489 amino acids) entered cilia (fig. S8, D to F), indicating that its C terminus targeted the kinase domain to cilia. Thus, the *dyf-5CA* mRNA confers substantial defects in DYF-5 protein synthesis and, at least, its C-terminal ciliary localization domain, which prevents the kinase from phosphorylating its ciliary substrates.

Nonsense-mediated dyf-5CA *mRNA decay*

We examined the fate of *dyf-5CA* mRNAs. The A-to-I RNA editing did not create a stop codon; however, the intron-retaining transcripts often contained a premature termination codon (PTC) (fig. S7E) and might be targeted for degradation by the nonsense-mediated mRNA decay (NMD) pathway (*24*). Therefore, we crossed *dyf-5CA* animals with NMD mutants, including *smg-1*, a critical kinase that mediates the phosphorylation of an RNA helicase UPF1, and *smg-3*, an ortholog of human UPF2 (*25*, *26*). The *dyf-5CA* transcripts in the *dyf-5CA; smg-1* double mutants became ~1.8-fold more abundant than in *dyf-5CA*, which was confirmed by using quantitative polymerase chain reaction (PCR) experiments (Fig. 3A and fig. S9A). *smg-1* or *smg-3* mutation, which did not affect WT cilia ($n > 100$), partially rescued the ciliary dye-filling defects of *dyf-5CA* (fig. S9B) and reduced IFT-particle aggregation (fig. S9, C to E), although *smg-1* mutation did not shorten the elongated cilia (fig. S9F). Blocking of NMD rescued the phenotype in Ullrich's disease by up-regulating the mRNA level of the mutant collagen subunit and forming a partially functional extracellular matrix (*27*). Likewise, the inhibition of NMD increases the *dyf-5CA* mRNA level, a small portion of which may translate into functional kinase, partially restoring *dyf-5CA* cilia.

RNA editing restricted two other ciliary kinases

To explore whether RNA editing restricts other kinases, we screened five kinases by overexpressing the kinase gDNA in WT. Two ciliary kinases, NEKL-4/NEK10 (never in mitosis kinase like) and DYF-18/CCRK (ortholog of human cyclin-dependent kinase 20) (*28*–*30*), when overexpressed, resulted in ciliary defects, aberrant antisense transcripts, and RNA editing on kinase pre-mRNA (Fig. 4, D to G, and fig. S10, A to D). *adr-2* deletion rescued the ciliary phenotypes in kinase overexpression (OE) animals (Fig. 4, D and E, and fig. S10, A and B), indicating that ciliary defects require RNA editing. However, we did not detect abnormalities in animal phenotypes, antisense transcripts, or RNA editing of the rest of the three cytoplasmic kinase OE strains, including the POLO kinase PLK-1, the AKT kinase family kinase AKT-1, and the adenosine monophosphate–activated protein kinase AAK-2 (fig. S11), suggesting that RNA editing may specifically regulate ciliary kinases.

Kinase hyperactivation caused aberrant RNA editing

We determined whether kinase hyperactivity or transgene overexpression caused ectopic RNA editing. Expression of a catalytic dead D609A mutation of NEKL-4 (*31*) did not disrupt cilia (Fig. 4, D and E) and did not generate aberrant antisense transcription or RNA editing (Fig. 4, F and G). Similarly, the catalytically dead DYF-18(D145A) OE did not show ciliary phenotypes (fig. S10, A and B). We detected antisense transcription and RNA editing at levels similar to those of DYF-18 OE animals (fig. S10, C and D). We speculate that DYF-18(D145A) overexpression might have a residual activity in vivo, triggering RNA editing, but could not impair cilia. Consistently, the same amount of gDNA transformation disrupted cilia in 82% of NEKL-4OE but in only 23% of DYF-18OE animals (Fig. 4, D and E, and fig. S10, A and B), which suggests that the ciliary structure might be more sensitive to up-regulated NEKL-4. Furthermore, the catalytically dead DYF-5(D150A) or DYF-5(K40A) (K40A, $Lys^{40} \rightarrow Ala$) (*32*) OE reduced the ciliary phenotype penetrance (fig. S12A), accompanying a reduction of aberrant antisense transcripts and RNA editing (fig. S12, B and C). Thus, kinase activity is a substantial cause for ciliary defects and abnormal RNA editing.

To overcome the limitation of using kinase overexpression as a proxy kinase hyperactivity, we introduced D150A in *dyf-5CA* animals (fig. S2A). DYF-5CA(D150A) protein was translated and correctly localized to cilia (fig. S12D), and *dyf-5CA(D150A)* animals did not generate any antisense transcript or RNA editing on *dyf-5* (Figs. 2B and 3, A and B), which indicated that aberrant antisense expression and RNA editing resulted from DYF-5CA hyperactivity rather than overexpression. Because of catalytic null, *dyf-5CA(D150A)* animals developed ectopically long cilia like *dyf-5* null (fig. S12D). These results support a direct contribution of kinase hyperactivity to aberrant antisense transcription and RNA editing. We suggest that transgene overexpression increased WT kinase levels, which up-regulates kinase activity, which in turn triggers antisense transcription and RNA editing of the native kinase locus.

Considering that ADR-2 is known to suppress transgene silencing through editing dsRNA that is produced from aberrant overlapping antisense transcription (*33*), ciliary phenotypes from kinase gDNA overexpression may involve transgene silencing. We introduced the *rde-1 (ne219)* and *rde-4(ne301)* mutant alleles defective in transgene silencing (*34*) into *dyf-5*, *nekl-4*, and *dyf-18* gDNA OE animals, but their ciliary phenotypes were not changed (fig. S12E), suggesting that transgene silencing may not be involved.

Transcription regulation of dyf-5CA *mRNA*

dyf-5CA animals have more RNA-seq reads in the first seven exons but fewer reads in the final four exons than that of WT (Fig. 3A). We compared the stability of the *dyf-5* transcripts in WT and *dyf-5CA* animals after treatment

with actinomycin D, which blocks transcription. *dyf-5* pre-mRNAs were degraded at a similar rate in both strains (fig. S13A), which suggests that DYF-5CA may not affect transcript stability. *dyf-5CA* animals abnormally transcribe antisense RNA from the fourth exon to the seventh intron of *dyf-5*; the resultant dsRNA may inhibit RNA polymerase II (Pol II) elongation (*35*). To assess the potential transcriptional changes, we performed RNA Pol II chromatin immunoprecipitation followed by sequencing chromatin immunoprecipitation–sequencing (ChIP-seq). Compared with WT, *dyf-5CA* animals increased Pol II occupancy in the third to eighth exon of *dyf-5* (fig. S13B). Similar increases were observed for the elongation-specific Pol II phosphorylated at serine 2 (fig. S13B). ChIP-seq of histone H3 trimethylation at Lys^{36} (H3K36me3) revealed increased H3K36me3 modifications in this region due to Pol II elongation and cotranscriptional deposition of H3K36me3 (fig. S13B). The increased densities indicate aberrantly high transcript elongation in the region and suggest that the lack of transcripts in the final few exons may result from transcription regulation. The *adr-2* deletion did not affect these densities in *dyf-5CA* (fig. S13B), which suggests that the increased occupancy may occur upstream or independent of RNA editing. By contrast, introducing D150A to *dyf-5CA* eliminates all the increased occupancy (fig. S13B), indicating that kinase hyperactivity causes abnormal transcription regulation.

Ciliary phenotypes in dyf-5CA *and* dyf-5 gOE *animals*

Why do *dyf-5CA* animals aberrantly elongate cilia but *dyf-5 gOE* worms shorten cilia? The absence of GFP fluorescence from *gfp::dyf-5CA* animals indicates that DYF-5CA protein is not produced, phenocopying the long cilia phenotype in *dyf-5* null. The *wrmScarlet::dyf-5 gOE* animals produced the wrmScarlet::DYF-5 fluorescence that was incorrectly distributed around the ciliary base (fig. S14, A, B, and G). Introducing *adr-2* null into *wrmScarlet::dyf-5 gOE* animals restored DYF-5 localization to the distal ciliary segments (fig. S14B) where DYF-5 is normally found, indicating that DYF-5 mislocalization owing to *dyf-5 gOE* is suppressed by the loss of ADR-2. Consistently, deletion of *adr-2* rescued the short cilia defects in *dyf-5 gOE* (fig. S14, C and G). Using quantitative fluorescence and quantitative real-time PCR experiments, we showed that the loss of ADR-2 does not affect the transgene expression level of *wrmScarlet::dyf-5 gOE* or *osm-6::gfpOE* (fig. S14, D and E), which indicates that ADR-2 does not regulate *dyf-5 gOE* phenotypes through promoting transgene expression. Because DYF-5 prevents cilia overgrowth (*2*), mislocalized DYF-5 at the ciliary base might have blocked cilia elongation in *dyf-5 gOE*. To test this possibility, we artificially targeted DYF-5 to the ciliary base by fusing *dyf-5* cDNA with a ciliary transition zone gene *mksr-2*. Indeed, *mksr-2::dyf-5cDNA OE* animals developed short cilia like that of *dyf-5 gOE* (fig. S14, C, F, and G); however, overexpressed DYF-5 protein from *dyf-5* cDNA correctly localized at the ciliary distal and did not impair cilia (fig. S14, B, C, and G). We suggest that as long as DYF-5 localizes at the correct place, it will not generate deleterious effects, even though its amount or activity might be higher than in WT animals. In *adr-2* mutants that have increased *dyf-5CA* levels but lose RNA editing, the full-length DYF-5CA proteins with the C-terminal localization domain can be produced and properly localize at the distal ciliary segments; therefore, no shortening of cilia was observed. We speculate that *dyf-5 gOE* animals expressed abundant *dyf-5* pre-mRNA, a portion of which, after RNA editing, may translate the DYF-5 kinase domain, incorrectly localizing around the ciliary base to inhibit cilia growth. By contrast, *dyf-5CA* animals expressed a much lower level of *dyf-5CA* pre-mRNA than *dyf-5 gOE*, and RNA editing may severely block DYF-5CA protein production, causing aberrant cilia elongation. If the *dyf-5CA* pre-mRNA increases under certain conditions, the animal may produce truncated DYF-5 protein that localizes at the ciliary base to inhibit cilia elongation.

Cilia-to-nuclei signaling

dyf-5CA animals exhibited higher levels of *dyf-5* mRNA than WT (Fig. 3A), which suggests negative transcriptional feedback so that the loss of DYF-5 up-regulates *dyf-5* expression. We found an up-regulation of other ciliary gene expressions in the *dyf-5CA* and *dyf-5 (mn400)* null allele and ciliary mutants that were defective in IFT-dynein CHE-3 and IFT-particle subunit DAF-10 (fig. S15 and data files S3 and S4). Flagella removal in *Chlamydomonas* caused a rapid transcription up-regulation of hundreds of flagellar-associated genes (*36*). Thus, a cell appears to sense whether cilia function or exist, and the "stressed" cilia may signal to the nucleus, requesting more ciliary proteins, which resembles the unfolded protein response (UPR) between the endoplasmic reticulum and the nucleus (*37*). Analogous to UPR regulation, an overphosphorylated ciliary protein of yet unknown identity may travel from *dyf-5CA* cilia into the nucleus to activate antisense transcription. This study describes a negative feedback loop in *dyf-5CA* animals. Given that cilia-to-nucleus signaling occurs in other ciliary mutants and deflagellated conditions, we speculate that the feedback loop elements may play roles in conditions beyond *dyf-5CA*. For example, WT worms that carry a mutation disabling the feedback loop (such as the intron substitution *dyf-5* allele in fig. S2A) might display phenotypes under certain stressful conditions.

Discussion

We propose a model through which RNA editing restricts the hyperactive DYF-5CA ciliary kinase (Fig. 4H): (i) In response to kinase hyperactivity, antisense RNAs are transcribed in nuclei and pair with the kinase pre-mRNA. (ii) The dsRNA recruits RNA deaminase to edit kinase pre-mRNA, which changes the kinase protein sequence and impairs kinase pre-mRNA splicing. (iii) The resultant IRs generate premature stop codons, which inhibit kinase translation and activate nonsense-mediated kinase mRNA decay. (iv) The loss of kinase production converts a biochemically hyperactive kinase to loss of function in vivo, phenocopying the kinase-null cilia. Considering that *dyf-5 gOE* generated more RNA editing sites on *dyf-5* locus than *dyf-5CA* (Fig. 2B), we suggest that kinase gRNA OE might involve additional regulation. However, the loss of RNA editing rescued ciliary defects under kinase gDNA OE conditions (Figs. 3C and 4, D and E, and fig. S10, A and B), underscoring the crucial role of RNA editing in safeguarding against the overactivation of kinases.

In addition to ciliary kinases, an increasing number of kinase mRNAs are hyperedited during disease progression. Examples include the protein kinase R during type I interferon responses and the CDK12 kinase mRNA from ovarian cancer cells (*38–40*). Although the mechanisms and impacts may differ from each of these RNAs, a parallel does emerge: RNA editing is linked to kinase hyperactivity. Recognizing this common linkage, we suggest that in response to physiological or pathological stimuli, RNA editing may target the kinase pre-mRNA, restricting kinase production and down-regulating its activity in vivo.

Our demonstration, in which suppression of RNA editing rescues ciliary defects caused by hyperactive kinases, connects cilia with other processes. Most of the ciliopathy genes are not ideal for drug development (*6*). We anticipate that the development of small molecules that target the pathways outside of cilia holds promise to devise strategies for the treatment of ciliopathies.

REFERENCES AND NOTES

1. P. Lahiry, A. Torkamani, N. J. Schork, R. A. Hegele, *Nat. Rev. Genet.* **11**, 60–74 (2010).
2. J. Burghoorn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7157–7162 (2007).
3. N. Vasan, J. Baselga, D. M. Hyman, *Nature* **575**, 299–309 (2019).
4. S. A. Berman, N. F. Wilson, N. A. Haas, P. A. Lefebvre, *Curr. Biol.* **13**, 1145–1149 (2003).
5. Y. Omori *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 22671–22676 (2010).
6. J. F. Reiter, M. R. Leroux, *Nat. Rev. Mol. Cell Biol.* **18**, 533–547 (2017).
7. R. K. Ozgül *et al.*, *Am. J. Hum. Genet.* **89**, 253–264 (2011).

8. B. A. Tucker *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, E569–E576 (2011).
9. P. D. Mace *et al*., *Nat. Commun.* **4**, 1681 (2013).
10. Z. Fu *et al*., *Mol. Cell. Biol.* **25**, 6047–6064 (2005).
11. L. A. Perkins, E. M. Hedgecock, J. N. Thomson, J. G. Culotti, *Dev. Biol.* **117**, 456–487 (1986).
12. G. Ou, O. E. Blacque, J. J. Snow, M. R. Leroux, J. M. Scholey, *Nature* **436**, 583–587 (2005).
13. K. Nishikura, *Nat. Rev. Mol. Cell Biol.* **17**, 83–96 (2016).
14. L. A. Tonkin *et al*., *EMBO J.* **21**, 6025–6035 (2002).
15. M. C. Washburn *et al*., *Cell Rep.* **6**, 599–607 (2014).
16. J. A. Arribere, H. Kuroyanagi, H. A. Hundley, *Genetics* **215**, 531–568 (2020).
17. H. Q. Zhao *et al*., *Genome Res.* **25**, 66–75 (2015).
18. H. Ohta, M. Fujiwara, Y. Ohshima, T. Ishihara, *Genetics* **180**, 785–796 (2008).
19. D. P. Reich, K. M. Tyc, B. L. Bass, *Genes Dev.* **32**, 271–282 (2018).
20. M. B. Warf, B. A. Shepherd, W. E. Johnson, B. L. Bass, *Genome Res.* **22**, 1488–1498 (2012).
21. F. Zhang, Y. Lu, S. Yan, Q. Xing, W. Tian, *Bioinformatics* **33**, 3538–3548 (2017).
22. R. Middleton *et al*., *Genome Biol.* **18**, 51 (2017).
23. J. Rodriguez, J. S. Menet, M. Rosbash, *Mol. Cell* **47**, 27–37 (2012).
24. T. Kurosaki, M. W. Popp, L. E. Maquat, *Nat. Rev. Mol. Cell Biol.* **20**, 406–420 (2019).
25. A. Grimson, S. O'Connor, C. L. Newman, P. Anderson, *Mol. Cell. Biol.* **24**, 7483–7490 (2004).
26. L. Johns, A. Grimson, S. L. Kuchma, C. L. Newman, P. Anderson, *Mol. Cell. Biol.* **27**, 5630–5638 (2007).
27. F. Usuki *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15037–15042 (2013).
28. R. R. Chivukula *et al*., *Nat. Med.* **26**, 244–251 (2020).
29. A. K. Maurya, T. Rogers, P. Sengupta, *Curr. Biol.* **29**, 1286–1300.e4 (2019).
30. K. M. Power *et al*., *PLOS Genet.* **16**, e1009052 (2020).
31. S. H. Schmidt *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14979–14988 (2019).
32. L. Xia *et al*., *J. Biol. Chem.* **277**, 35422–35433 (2002).
33. S. W. Knight, B. L. Bass, *Mol. Cell* **10**, 809–817 (2002).
34. H. Tabara, E. Yigit, H. Siomi, C. C. Mello, *Cell* **109**, 861–871 (2002).
35. V. Pelechano, L. M. Steinmetz, *Nat. Rev. Genet.* **14**, 880–893 (2013).
36. A. J. Albee *et al*., *G3* **3**, 979–991 (2013).
37. P. Walter, D. Ron, *Science* **334**, 1081–1086 (2011).
38. C. Calabrese *et al*., *Nature* **578**, 129–136 (2020).
39. H. Chung *et al*., *Cell* **172**, 811–824.e14 (2018).
40. J. J. Ishizuka *et al*., *Nature* **565**, 43–48 (2019).

ACKNOWLEDGMENTS

We thank J. Reiter, X. Yang, V. Ambros, D. Xue, Z. Lu, X. Shen, Y. Qi, S. Cai, and Y. Wang for discussion. **Funding:** This work was supported by the following funding programs: National Natural Science Foundation of China (grants 31991190, 31730052, 31525015, 31861143042, 31561130153, 31671444, and 31871352) and National Key R&D Program of China (2017YFA0503501, 2019YFA0508401, and 2017YFA0102900). **Author contributions:** D.L., P.Y., Z.Z., W.L., Q.C.Z., J.B.L., and G.O. conceived and designed the project. D.L., Z.Z., and P.Y. performed research; D.L. and Y.L. analyzed data; D.L., Z.Z., W.L., Q.C.Z., J.B.L., and G.O. wrote the paper. **Competing interests:** G.O., W.L., D.L., Z.Z., and P.Y. are inventors on patent application no. CN202110893049.4, submitted by Tsinghua University, that covers applications of RNA editing on genetic disorders that are associated with gain-of-function proteins. **Data and materials availability:** All data are available in the main text or supplementary materials. All sequencing data from this study have been submitted to the NCBI Sequence Read Archive (SRA; www.ncbi.nlm.nih.gov/sra) with BioProject accession number PRJNA655606.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/984/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S5
References (*41*–*47*)
MDAR Reproducibility Checklist
Data S1 to S4

24 August 2020; resubmitted 1 April 2021
Accepted 16 July 2021
10.1126/science.abd8971

CORONAVIRUS

Chimeric spike mRNA vaccines protect against Sarbecovirus challenge in mice

David R. Martinez[1]*, Alexandra Schäfer[1], Sarah R. Leist[1], Gabriela De la Cruz[2], Ande West[1], Elena N. Atochina-Vasserman[3], Lisa C. Lindesmith[1], Norbert Pardi[3], Robert Parks[4], Maggie Barr[4], Dapeng Li[4], Boyd Yount[1], Kevin O. Saunders[4], Drew Weissman[3], Barton F. Haynes[4], Stephanie A. Montgomery[5], Ralph S. Baric[1]*

The emergence of severe acute respiratory syndrome coronavirus (SARS-CoV) in 2003 and SARS-CoV-2 in 2019 highlights the need to develop universal vaccination strategies against the broader *Sarbecovirus* subgenus. Using chimeric spike designs, we demonstrate protection against challenge from SARS-CoV, SARS-CoV-2, SARS-CoV-2 B.1.351, bat CoV (Bt-CoV) RsSHC014, and a heterologous Bt-CoV WIV-1 in vulnerable aged mice. Chimeric spike messenger RNAs (mRNAs) induced high levels of broadly protective neutralizing antibodies against high-risk Sarbecoviruses. By contrast, SARS-CoV-2 mRNA vaccination not only showed a marked reduction in neutralizing titers against heterologous Sarbecoviruses, but SARS-CoV and WIV-1 challenge in mice resulted in breakthrough infections. Chimeric spike mRNA vaccines efficiently neutralized D614G, mink cluster five, and the UK B.1.1.7 and South African B.1.351 variants of concern. Thus, multiplexed-chimeric spikes can prevent SARS-like zoonotic coronavirus infections with pandemic potential.

A novel severe acute respiratory syndrome coronavirus (SARS-CoV) emerged in 2003 and caused more than 8000 infections and ~800 deaths worldwide (*1*). In 2012, the Middle East respiratory syndrome coronavirus (MERS-CoV) emerged in Saudi Arabia (*2*), with multiple outbreaks that have resulted in at least ~2600 cases and 900 deaths (*3*). In December 2019, another novel human SARS-like virus from the genus *Betacoronavirus* and subgenus *Sarbecovirus* emerged in Wuhan China, designated SARS-CoV-2, causing the ongoing COVID-19 pandemic (*4*, *5*).

Bats are known reservoirs of SARS-like coronaviruses (CoVs) and harbor high-risk "preemergent" SARS-like variant strains, such as WIV-1-CoV and RsSHC014-CoV, which are able to use human ACE2 (angiotensin-converting enzyme 2) receptors for entry, replicate efficiently in human primary airway epithelial cells, and may escape existing countermeasures (*6*, *7*). Given the high pandemic potential of zoonotic and epidemic Sarbecoviruses (*8*), the development of countermeasures such as broadly effective vaccines, antibodies, and drugs is a global health priority (*9*–*11*).

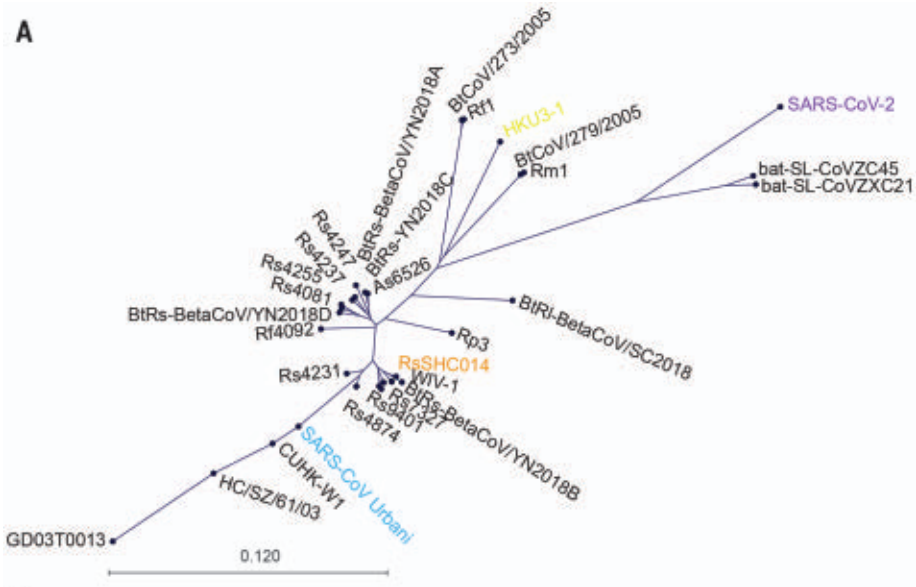
[1]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [2]Lineberger Comprehensive Cancer Center, University of North Carolina School of Medicine, Chapel Hill, NC, USA. [3]Infectious Disease Division, Department of Medicine, Perelman School of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [4]Duke Human Vaccine Institute, Duke University School of Medicine, Durham, NC, USA. [5]Department of Laboratory Medicine and Pathology, University of North Carolina School of Medicine, Chapel Hill, NC, USA.
***Corresponding author. Email: david.rafael.martinez@gmail.com (D.R.M.); rbaric@email.unc.edu (R.S.B.)**

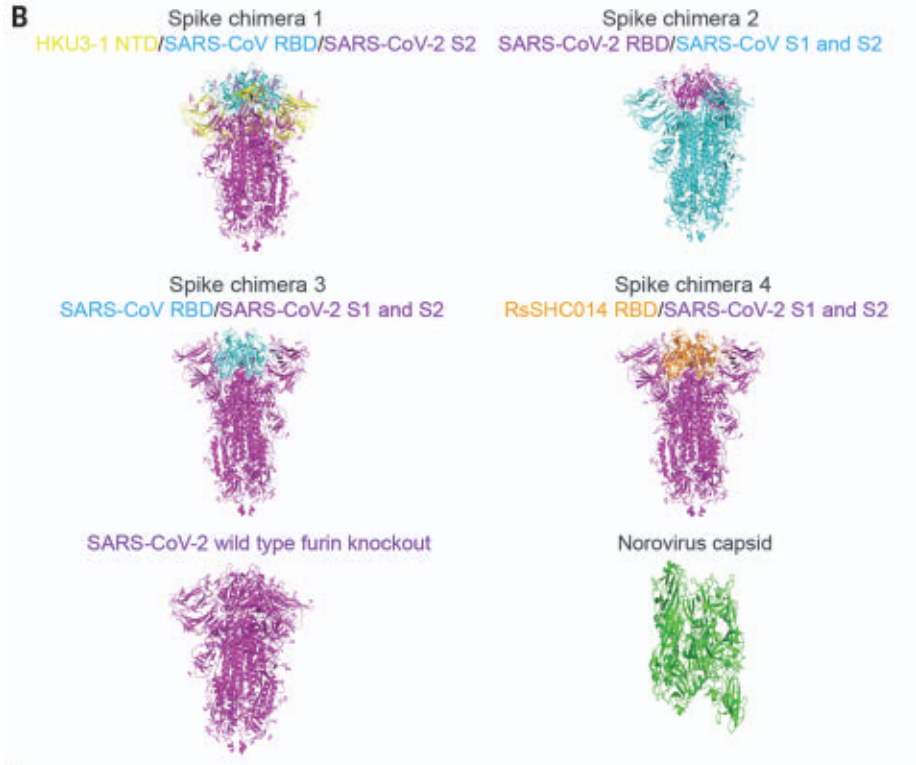
Sarbecovirus spike proteins have immunogenic domains: the receptor binding domain (RBD), the N-terminal domain (NTD), and the subunit 2 (S2) (*12*, *13*). RBD, NTD, and to a lesser extent S2 are targets for potent neutralizing and non-neutralizing antibodies elicited to SARS-CoV-2 and MERS-CoV spike (*12*, *14*–*19*). Passive immunization with SARS-CoV-2 NTD-specific antibodies protect naïve mice from challenge, demonstrating that the NTD is a target of protective immunity (*12*, *19*, *20*). However, it remains unclear whether vaccine-elicited neutralizing antibodies can protect against in vivo challenge with heterologous epidemic and bat coronaviruses. We generated nucleoside-modified mRNA-lipid nanoparticle (LNP) vaccines expressing chimeric spikes that contain admixtures of different RBD, NTD, and S2 modular domains from zoonotic, epidemic, and pandemic CoVs and examined their efficacy against homologous and heterologous Sarbecovirus challenge in aged mice.

Results

Design and expression of chimeric spike constructs to cover pandemic and zoonotic SARS-related coronaviruses

Sarbecoviruses exhibit considerable genetic diversity (Fig. 1A), and SARS-like bat CoVs (Bt-CoVs) are recognized threats to human health (*6*, *8*). Because potent neutralizing antibody epitopes exist in each of the modular structures on CoV spikes (*21*), we hypothesized that chimeric spikes that encode NTD, RBD, and S2 domains into "bivalent" and "trivalent" vaccine immunogens have the potential to elicit broad protective antibody responses against clades I to III Sarbecoviruses. We designed four sets of chimeric spikes.





C

Chimeric spike mRNA-LNP immunogens

Chimera	RBD	NTD	S2
1	SARS-CoV	HKU3-1	SARS-CoV-2
2	SARS-CoV-2	SARS-CoV	SARS-CoV
3	SARS-CoV	SARS-CoV-2	SARS-CoV-2
4	RsSHC014	SARS-CoV-2	SARS-CoV-2

Fig. 1. Genetic design of chimeric Sarbecovirus spike vaccines. (**A**) Genetic diversity of pandemic and bat zoonotic coronaviruses. HKU3-1 is shown in yellow, SARS-CoV is shown in light blue, RsSHC014 is shown in orange, and SARS-CoV-2 is shown in purple. (**B**) Spike chimera 1 includes the NTD from HKU3-1, the RBD from SARS-CoV, and the rest of the spike from SARS-CoV-2. Spike chimera 2 includes the RBD from SARS-CoV-2 and the NTD and S2 from SARS-CoV. Spike chimera 3 includes the RBD from SARS-CoV and the NTD and S2 SARS-CoV-2. Spike chimera 4 includes the RBD from RsSHC014 and the rest of the spike from SARS-CoV-2. SARS-CoV-2 furin KO spike vaccine and is the norovirus capsid vaccine. (**C**) Table summary of chimeric spike constructs.

Chimera 1 included the NTD from clade II Bt-CoV Hong Kong University 3-1 (HKU3-1), the clade I SARS-CoV RBD, and the clade III SARS-CoV-2 S2 (Fig. 1B). Chimera 2 included SARS-CoV-2 RBD and SARS-CoV NTD and S2 domains (*11*). Chimera 3 included the SARS-CoV RBD and SARS-CoV-2 NTD and S2, whereas chimera 4 included the RsSHC014 RBD and SARS-CoV-2 NTD and S2. We also generated a monovalent SARS-CoV-2 spike furin knockout (KO) vaccine, partially phenocopying the Moderna and Pfizer mRNA vaccines in human use, and a negative control norovirus GII capsid vaccine (Fig. 1, B and C). We generated these chimeric spikes and control spikes as lipid nanoparticle-encapsulated, nucleoside-modified mRNA vaccines with LNP adjuvants (mRNA-LNP), as described previously (*22*). This mRNA LNP stimulates robust T follicular helper cell activity, germinal center B cell responses, durable long-lived plasma cells, and memory B cell responses (*23*, *24*). We verified their chimeric spike expression in human embryonic kidney (HEK) cells (fig. S1B). To confirm that scrambled coronavirus spikes are biologically functional, we also designed and recovered several high-titer recombinant live viruses of RsSHC014/SARS-CoV-2 NTD, RBD, and S2 domain chimeras that included deletions in nonessential, accessory open reading frame 7 (ORF7) and ORF8 that encoded nanoluciferase (fig. S1C). SARS-CoV-2 ORF7 and -8 antagonize innate immune signaling pathways (*25*, *26*), and deletions in these ORFs are associated with attenuated disease in humans (*27*, *28*).

Immunogenicity of mRNAs expressing chimeric spike constructs against coronaviruses

We next sought to determine whether simultaneous immunization with mRNA-LNP expressing the chimeric spikes of diverse Sarbecoviruses was a feasible strategy to elicit broad binding and neutralizing antibodies. We immunized aged mice with the chimeric spikes formulated to induce cross-reactive responses against multiple divergent clades I to III Sarbecoviruses, a SARS-CoV-2 furin KO spike, and a GII.4 norovirus capsid negative control. Group 1 was primed and boosted with chimeric spikes 1, 2, 3, and 4 (fig. S1A). Group 2 was primed with chimeric spikes 1 and 2 and boosted with chimeric spikes 3 and 4 (fig. S1A). Group 3 was primed and boosted with chimeric spike 4 (fig. S1A). Group 4 was primed and boosted with the monovalent SARS-CoV-2 furin KO spike (fig. S1A). Last, group 5 was primed and boosted with a norovirus capsid GII.4 Sydney 2011 strain (fig. S1A). We then examined the binding antibody responses by means of enzyme-linked immunosorbent assay (ELISA) against a diverse panel of CoV spike proteins that included epidemic, pandemic, and zoonotic coronaviruses.

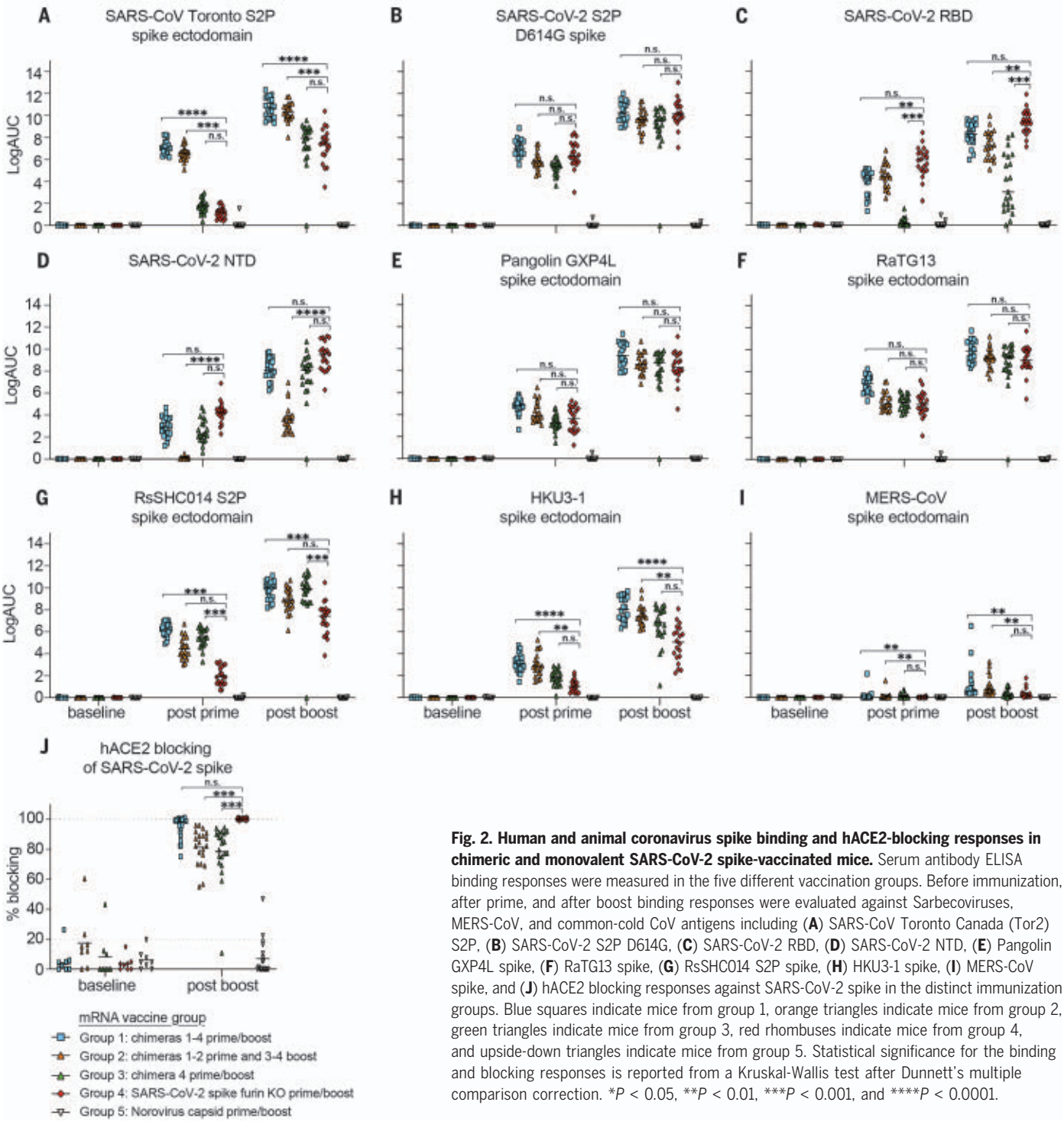


Fig. 2. Human and animal coronavirus spike binding and hACE2-blocking responses in chimeric and monovalent SARS-CoV-2 spike-vaccinated mice. Serum antibody ELISA binding responses were measured in the five different vaccination groups. Before immunization, after prime, and after boost binding responses were evaluated against Sarbecoviruses, MERS-CoV, and common-cold CoV antigens including (**A**) SARS-CoV Toronto Canada (Tor2) S2P, (**B**) SARS-CoV-2 S2P D614G, (**C**) SARS-CoV-2 RBD, (**D**) SARS-CoV-2 NTD, (**E**) Pangolin GXP4L spike, (**F**) RaTG13 spike, (**G**) RsSHC014 S2P spike, (**H**) HKU3-1 spike, (**I**) MERS-CoV spike, and (**J**) hACE2 blocking responses against SARS-CoV-2 spike in the distinct immunization groups. Blue squares indicate mice from group 1, orange triangles indicate mice from group 2, green triangles indicate mice from group 3, red rhombuses indicate mice from group 4, and upside-down triangles indicate mice from group 5. Statistical significance for the binding and blocking responses is reported from a Kruskal-Wallis test after Dunnett's multiple comparison correction. *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001.

Mice in groups 1 and 2 generated the highest-magnitude responses to SARS-CoV Toronto Canada isolate (Tor2), RsSHC014, and HKU3-1 spike as compared with group 4 (Fig. 2, A, G, and H). Whereas mice in group 2 generated lower-magnitude binding responses to both SARS-CoV-2 RBD (Fig. 2C) and SARS-CoV-2 NTD (Fig. 2D), mice in group 1 generated similar-magnitude binding antibodies to SARS-CoV-2 D614G (in which aspartic acid at position 614 is replaced with glycine) as compared with that of mice immunized with the SARS-CoV-2 furin KO spike mRNA-LNP (Fig. 2B). Mice in groups 1 and 2 generated similar-magnitude binding antibody responses against SARS-CoV-2 D614G, Pangolin GXP4L, and RaTG13 spikes (Fig. 2, B, E, and F) compared with those of mice from group 4. Mice in group 1 and group 4 elicited high-magnitude levels of hACE2 blocking responses, as compared with those of groups 2 and 3 (Fig. 2J). Because binding antibody responses after boost mirrored the trend of the after-prime responses, it is likely that the second dose is boosting immunity to the vaccine antigens in the prime (Fig. 2). Last, we did not observe cross-binding antibodies against common-cold CoV spike

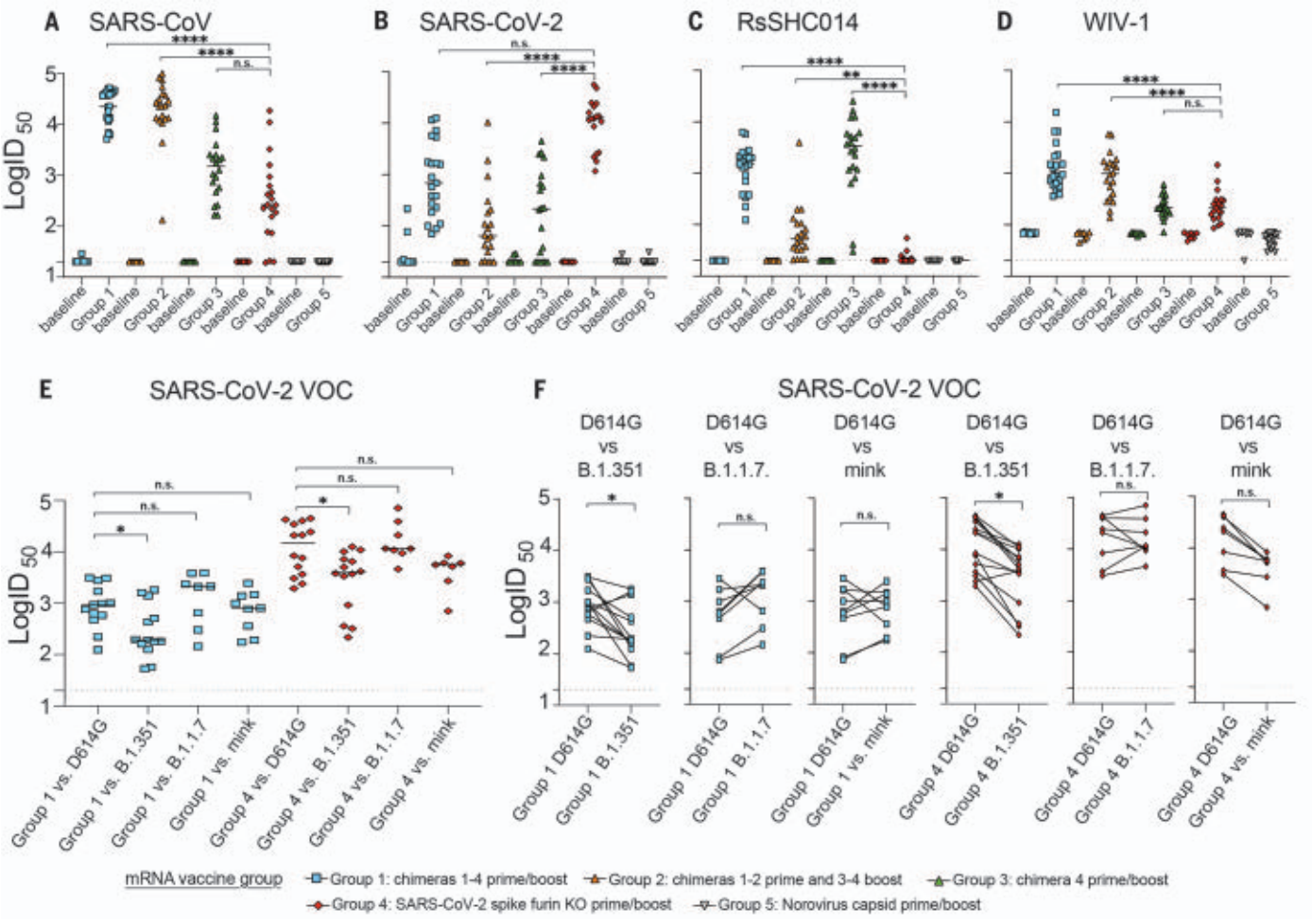


Fig. 3. Live Sarbecovirus neutralizing antibody responses in vaccinated mice. Neutralizing antibody responses in mice from the five different vaccination groups were measured by using nanoluciferase-expressing recombinant viruses. (**A**) SARS-CoV neutralizing antibody responses from baseline and after boost in the distinct vaccine groups. (**B**) SARS-CoV-2 neutralizing antibody responses from baseline and after boost. (**C**) RsSHC014 neutralizing antibody responses from baseline and after boost. (**D**) WIV-1 neutralizing antibody responses from baseline and after boost. (**E**) The neutralization activity in groups 1 and 4 against SARS-CoV-2 D614G, South African B.1.351, UK B.1.1.7, and mink cluster 5 variant. (**F**) Neutralization comparison of SARS-CoV-2 D614G versus South African B.1.351, versus UK B.1.1.7, and versus mink cluster 5 variant. Statistical significance for the live-virus neutralizing antibody responses is reported from a Kruskal-Wallis test after Dunnett's multiple comparison correction. *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001.

antigens from HCoV-HKU1, HCoV-NL63, and HCoV-229E in most of the vaccine groups (fig. S2, A to D), but we did observe low binding levels against more distant group 2C MERS-CoV (Fig. 2I) and other Betacoronaviruses such as group 2A HCoV-OC43 in vaccine groups 1 and 2 (fig. S2B). These results suggest that chimeric spike vaccines elicit broader and higher-magnitude binding responses against pandemic and bat SARS-like viruses as compared with those of monovalent SARS-CoV-2 spike vaccines.

Neutralizing antibody responses against live Sarbecoviruses and variants of concern

We then examined the neutralizing antibody responses against SARS-CoV, Bt-CoV RsSHC014, Bt-CoV WIV-1, and SARS-CoV-2 including variants of concern using live viruses as previously described (Fig. 3, A to D) (*17*). Group 4 SARS-CoV-2 S mRNA–vaccinated animals mounted a robust response against SARS-CoV-2; however, responses against SARS-CoV, RsSHC014, and WIV-1 were 18-, >300- or 116-fold decreased, respectively (Fig. 3, A to D, and fig. S3, G and H). By contrast, aged mice in group 2 showed a 42- and twofold increase in neutralizing titer against SARS-CoV and WIV1 and less than onefold decrease against RsSHC014 relative to SARS-CoV-2 neutralizing titers (Fig. 3, A to D, and fig. S3, C and D). Mice in group 3 elicited thee- and sevenfold increased neutralizing titers against SARS-CoV and RsSHC014 yet showed a threefold decrease in WIV-1 neutralizing titers relative to SARS-CoV-2 (Fig. 3, A to D, and fig. S3, E and F). Last, mice in group 1 generated the most balanced and highest neutralizing titers, which were 13- and 1.2-fold increased against SARS-CoV and WIV-1 and less than onefold decreased against RsSHC014 relative to the SARS-CoV-2 neutralizing titers (Fig. 3, A to D, and fig. S3, A and B). The serum of mice from groups 1 and 4 neutralized the dominant D614G variant with similar potency as that of the wild-type D614 nonpredominant variant, and both groups had similar neutralizing antibody responses against the UK B.1.1.7 and the mink cluster 5 variant as compared with the D614G variant (Fig. 3, E and F). Despite the significant but small reduction in neutralizing activity against the B.1.351 variant of concern (VOC), we did not observe a complete ablation in neutralizing activity in either group. Mice from groups 1 and 2 elicited lower binding and neutralizing responses to SARS-CoV-2 as compared with those of group 4, perhaps reflecting a decreased

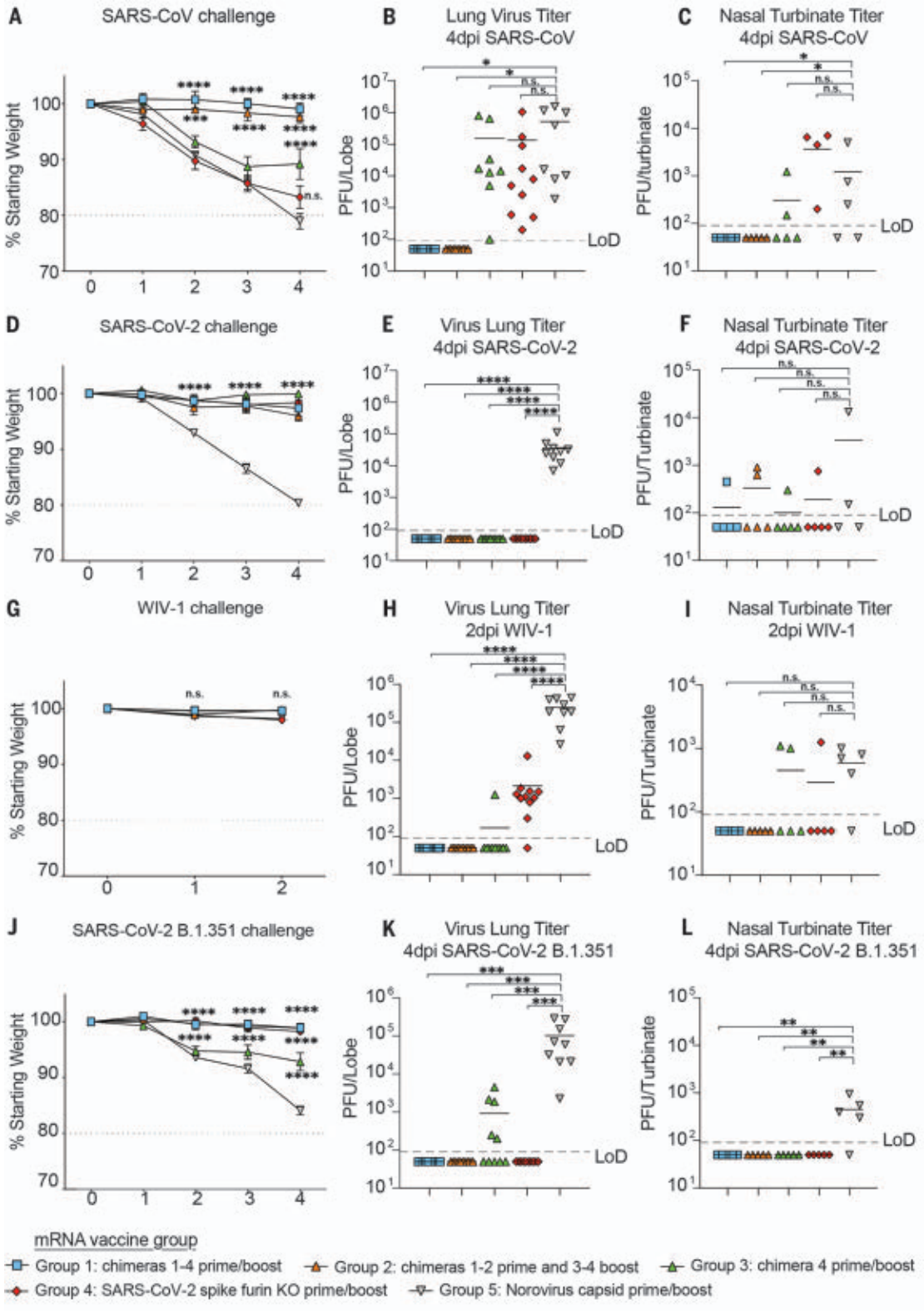


Fig. 4. In vivo protection against Sarbecovirus challenge after mRNA-LNP vaccination. (**A**) Percent starting weight from the different vaccine groups of mice challenged with SARS-CoV MA15. (**B**) SARS-CoV MA15 lung viral titers in mice from the distinct vaccine groups. (**C**) SARS-CoV MA15 nasal turbinate titers. (**D**) Percent starting weight from the different vaccine groups of mice challenged with SARS-CoV-2 MA10. (**E**) SARS-CoV-2 MA10 lung viral titers in mice from the distinct vaccine groups. (**F**) SARS-CoV-2 MA10 nasal turbinate titers. (**G**) Percent starting weight from the different vaccine groups of mice challenged with WIV-1. (**H**) WIV-1 lung viral titers in mice from the distinct vaccine groups. (**I**) WIV-1 nasal turbinate titers. (**J**) Percent starting weight from the different vaccine groups of mice challenged with SARS-CoV-2 B.1.351. (**K**) SARS-CoV-2 B.1.351 lung viral titers in mice from the distinct vaccine groups. (**L**) SARS-CoV-2 B.1.351 nasal turbinate titers. The vaccines used in the different groups are denoted at bottom. Statistical significance for weight loss is reported from a two-way analysis of variance (ANOVA) after Dunnett's multiple comparison correction. For lung and nasal turbinate titers, statistical significance is reported from a one-way ANOVA after Tukey's multiple comparison correction. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, and $****P < 0.0001$.

amount of mRNA vaccine incorporated into multiplexed formulations; the monovalent vaccines may drive a more focused B cell response to SARS-CoV-2, whereas chimeric spike antigens lead to more breadth against distant Sarbecoviruses. Thus, both monovalent SARS-CoV-2 vaccines and multiplexed chimeric spikes elicit neutralizing antibodies against newly emerged SARS-CoV-2 variants, and multiplexed chimeric spike vaccines outperform the monovalent SARS-CoV-2 vaccines in terms of breadth against multiclade Sarbecoviruses.

In vivo protection against heterologous Sarbecovirus challenge

To assess the ability of the mRNA-LNP vaccines to mediate protection against previously epidemic SARS-CoV, pandemic SARS-CoV-2, and Bt-CoVs, we challenged the different

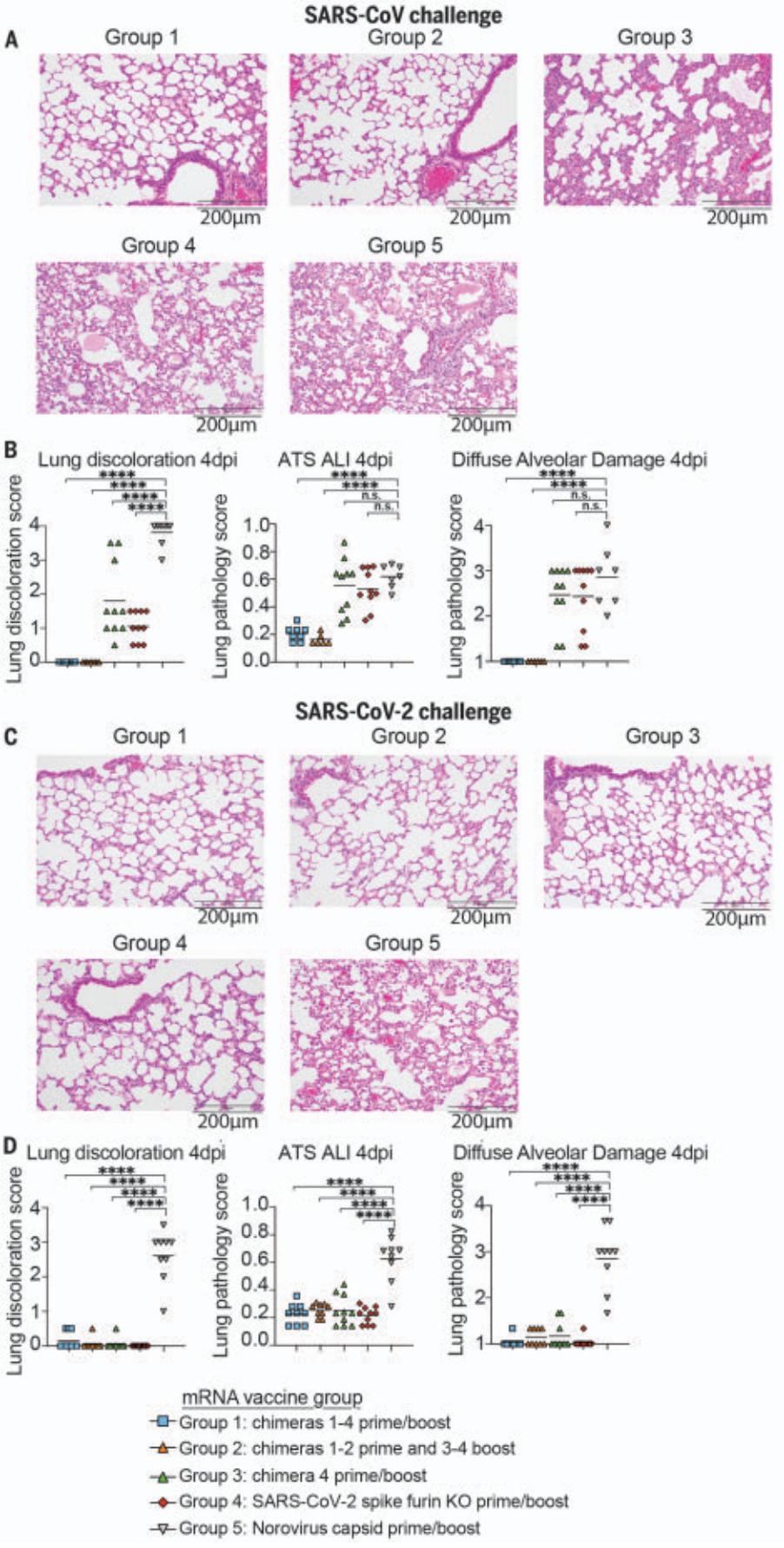


Fig. 5. Lung pathology in vaccinated mice after SARS-CoV and SARS-CoV-2 challenge. (**A**) Hematoxylin and eosin 4 days after infection lung analysis of SARS-CoV MA15–challenged mice from the different groups: group 1, chimeras 1 to 4 prime and boost; group 2, chimeras 1 and 2 prime and 3 and 4; group 3, chimera 4 prime and boost, SARS-CoV-2 furin KO prime and boost, and norovirus capsid prime and boost. (**B**) Lung pathology quantitation in SARS-CoV MA15–challenged mice from the different groups. Macroscopic lung discoloration score, microscopic ALI score, and DAD in day 4 after infection lung tissues are shown. (**C**) Hematoxylin and eosin 4 days after infection lung analysis of SARS-CoV-2 MA10–challenged mice from the different groups. (**D**) Lung pathology measurements in SARS-CoV-2 MA10–challenged mice from the different groups. Macroscopic lung discoloration score, microscopic ALI score, and DAD in day 4 after infection lung tissues are shown. Statistical significance is reported from a one-way ANOVA after Dunnet's multiple comparison correction. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, and $****P < 0.0001$.

groups and observed the mice for signs of clinical disease. Mice from group 1 or group 2 were completely protected from weight loss and lower- and upper-airway virus replication as measured with infectious virus plaque assays after 2003 SARS-CoV mouse-adapted (MA15) challenge (Fig. 4, A, B, and C). Similarly, these two vaccine groups were also protected against SARS-CoV-2 mouse-adapted (MA10) challenge. By contrast, group 3 showed some protection against SARS-CoV MA15–induced weight loss but not against viral replication in the lung or nasal turbinates. Group 3 was fully protected against SARS-CoV-2 MA10 challenge. By contrast, group 5 vaccinated mice developed severe disease, including mortality in both SARS-CoV MA15 and SARS-CoV-2 MA10 infections (fig. S5, B and C). Monovalent SARS-CoV-2 mRNA vaccines were highly efficacious against SARS-CoV-2 MA10 challenge but failed to protect against SARS-CoV MA15–induced weight loss and replication in the lower and upper respiratory tract (Fig. 4, A, B, and C), suggesting that SARS-CoV-2 mRNA-LNP vaccines are not likely to protect against future SARS-CoV emergence events. Mice from groups 1 to 4 were completely protected from weight loss and lower airway SARS-CoV-2 MA10 replication (Fig. 4, D, E, and F). Using both a Bt-CoV RsSHC014 full-length virus and a more virulent RsSHC014-MA15 chimera in mice (*6*), we also demonstrated protection in groups 1 to 3 against RsSHC014 replication in the lung and nasal turbinates (fig. S4) but not in mice that received the SARS-CoV-2 mRNA vaccine. Group 5 control mice challenged with RsSHC014-MA15 developed disease, including mortality (fig. S5D). Group 3 mice, which received a SARS-CoV-2 NTD/RsSHC014 RBD/SARS-CoV-2 S2, were fully protected against both SARS-CoV-2 and RsSHC014 challenge, whereas group 4 mice were not, demonstrating that a single NTD and RBD chimeric spike can protect against more than one virus compared with a monovalent spike.

We then performed a heterologous challenge experiment with the bat preemergent WIV-1-CoV (*7*). Mice from groups 1 and 2 were fully protected against heterologous WIV-1 challenge, whereas mice that received the SARS-CoV-2 mRNA vaccine had breakthrough replication in the lung (Fig. 4, G, H, and I). We also challenged with a virulent form of SARS-CoV-2 VOC B.1.351, which contains deletions in the NTD and mutations in the RBD, and observed full protection in vaccine groups 1, 2, and 4 compared with that in controls, whereas breakthrough replication was observed in group 3, further indicating the importance of the NTD in vaccine-mediated protection (Fig. 4, J, K, and L). The reduced protection against the B.1.351 variant containing NTD deletions indicates that the NTD is a clear target of protective immunity and that its inclusion in vaccination strategies, as opposed to RBD-alone vaccines, may be required to achieve full protection. Moreover, the SARS-CoV-2 mRNA vaccine protected against SARS-CoV-2 B.1.351 challenge in aged mice despite a reduction in the neutralizing activity against this VOC.

Lung pathology and cytokines in mRNA-LNP–vaccinated mice challenged with epidemic and pandemic coronaviruses

To quantify the pathological features of acute lung injury (ALI) in mice, we used a tool from the American Thoracic Society (ATS). We similarly scored lung tissue sections for diffuse alveolar damage (DAD), the pathological hallmark of ALI (*29*, *30*). We observed significant lung pathology with both the ATS and DAD scoring tools in groups 4 and 5 vaccinated animals. By contrast, multiplexed chimeric spike vaccine formulations in groups 1 and 2 provided complete protection from lung pathology after SARS-CoV MA15 challenge (Fig. 5, A and B). Mice immunized with the SARS-CoV-2 mRNA vaccine that showed breakthrough infection with SARS-CoV MA15 developed similar lung inflammation as that of control vaccinated animals, potentially suggesting that future outbreaks of SARS-CoV may cause disease even in individuals vaccinated with SARS-CoV-2. Because eosinophilic infiltrates have been observed in vaccinated, 2003 SARS-CoV–challenged mice previously (*31*), with immunohistochemistry we analyzed lung tissues in protected versus infected animals with SARS-CoV MA15 for eosinophilic infiltrates (fig. S6). Groups 1 and 2 contained rare, scattered eosinophils in the interstitium. Group 3 showed bronchus-associated lymphoid tissue. By contrast, group 4 and group 5 contained frequent perivascular cuffs with prevalent eosinophils. All groups challenged with SARS-CoV-2 MA10 were protected against lung pathology compared with the norovirus capsid-immunized control group, supporting the hypothesis that the SARS-CoV-2 NTD present in the chimeric spike from group 3 is sufficient for protection (Fig. 5, C and D).

We measured lung proinflammatory cytokines and chemokines in the different vaccination groups. Groups 1 and 2 had baseline levels of macrophage-activating cytokines and chemokines, including interleukin-6 (IL-6), chemokine (C-C motif) ligand 2 (CCL2), IL-1α, granulocyte colony-stimulating factor (G-CSF), and CCL4, compared with group 5 after SARS-CoV MA15 challenge (fig. S7A). Group 3 and group 4 showed high and indistinguishable levels of IL-6, CCL2, IL-1α, G-CSF, and CCL4 compared with those of group 5 mice after SARS-CoV MA15 challenge. After SARS-CoV-2 MA10 challenge, group 4 and group 1 showed the lowest levels of IL-6 and G-CSF relative to that in group 5 controls (fig. S7B), and we only observed significant reductions in CCL2, IL-1α, and CCL4 lung levels in groups 3 and 4 compared with the group 5 control, despite full protection from both weight loss and lower-airway viral replication.

Discussion

The Moderna and Pfizer/BioNTech SARS-CoV-2 mRNA-LNP vaccines were safe and efficacious against SARS-CoV-2 infections in large phase 3 efficacy human clinical trials (*32*–*34*), but there is a growing concern regarding VOCs such as South African B.1.351, which is five- to sixfold more resistant to vaccine-elicited polyclonal neutralizing antibodies (*35*). We sought to replicate the mRNA platform to formulate chimeric vaccines that specifically target distant Sarbecovirus strains. A caveat of including multiple chimeric spikes in a single shot is the potential formation of heterotrimers not present in the intended vaccine formulation. Chimera 4, which contains the RsSHC014 RBD and SARS-CoV-2 NTD and S2, elicited binding and neutralizing antibodies, and mice were fully protected from Bt-CoV RsSHC014 and SARS-CoV-2 challenge, whereas SARS-CoV-2 full length did not fully protect against RsSHC014, suggesting that CoV spike vaccines can be designed to maximize their display of protective epitopes and indicates that NTD/RBD/S2 chimeric spikes may enhance protection relative to monovalent spikes. Because the NTD, RBD, and S2 contain epitopes that are targeted by protective antibodies (*17*, *19*, *36*), modular chimeric spikes may provide a way to design CoV spikes to elicit protective immunity against three Sarbecoviruses as compared with a single Sarbecovirus by a monovalent spike. The lack of protection against WIV-1 and SARS-CoV and only partial protection against RsSHC014 challenge in SARS-CoV-2 immunized mice indicates the need for the development of universal vaccination strategies that can achieve broader coverage against preemergent bat SARS-CoV–like and SARS-CoV-2–like viruses. Despite the lower-magnitude antibody responses against SARS-CoV-2 in the chimeric spike groups, a clear advantage of our chimeric spike vaccines is the clear breadth of protection against multiclade Sarbecoviruses and SARS-CoV-2 variants compared with that from the monovalent SARS-CoV-2 vaccine. Although other strategies exist, including multiplexing mosaic Sarbecovirus RBDs (*37*) and RBDs on nanoparticles (*38*), chimeric spike mRNA-LNP vaccination can achieve broad protection by using existing manufacturing technologies and are portable to other high-risk emerging coronaviruses such as group 2C MERS-CoV–related strains. Thus, chimeric spikes can clearly protect against more than one Sarbecovirus, but it is possible that multiplexed full-length spikes may protect against Sarbecoviruses.

As previously reported with RNA recombinant viruses, live Sarbecoviruses lacking ORF7/

ORF8 but containing distinct SARS-CoV-2 antigenic domains were viable, reaffirming the known interchangeability and functional plasticity of the CoV spike (*21*, *39*, *40*). Our demonstration of cross-protection against multiple Sarbecovirus strains in mice lends support to the hypothesis that universal vaccines against group 2B CoVs are likely achievable. Moving forward, it will be important to determine whether other combinations of chimeric mRNA-LNP vaccines from other SARS-like viruses are protective, elicit broad T cell responses, prevent the rapid emergence of escape viruses, elicit protective responses in nonhuman primate models of Sarbecovirus pathogenesis, and can boost Sarbecovirus protective breadth in SARS-CoV-2–vaccinated or convalescent individuals.

REFERENCES AND NOTES

1. J. D. Cherry, P. Krogstad, *Pediatr. Res.* **56**, 1–5 (2004).
2. A. M. Zaki, S. van Boheemen, T. M. Bestebroer, A. D. Osterhaus, R. A. Fouchier, *N. Engl. J. Med.* **367**, 1814–1820 (2012).
3. C. I. Paules, H. D. Marston, A. S. Fauci, *JAMA* **323**, 707–708 (2020).
4. P. Zhou *et al.*, *Nature* **579**, 270–273 (2020).
5. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, *Nat. Microbiol.* **5**, 536–544 (2020).
6. V. D. Menachery *et al.*, *Nat. Med.* **21**, 1508–1513 (2015).
7. V. D. Menachery *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3048–3053 (2016).
8. B. Hu *et al.*, *PLOS Pathog.* **13**, e1006698 (2017).
9. T. P. Sheahan *et al.*, *Sci. Transl. Med.* **9**, eaal3653 (2017).
10. C. G. Rappazzo *et al.*, *Science* **371**, 823–829 (2021).
11. D. R. Martinez *et al.*, A broadly neutralizing antibody protects against SARS-CoV, pre-emergent bat CoVs, and SARS-CoV-2 variants in mice. bioRxiv [Preprint] 28 April 2021. doi:10.1101/2021.04.27.441655.
12. W. N. Voss *et al.*, *Science* **372**, 1108–1112 (2021).
13. C. Graham *et al.*, *Immunity* **54**, 1276–1289.e6 (2021).
14. L. Premkumar *et al.*, *Sci. Immunol.* **5**, eabc8413 (2020).
15. L. Liu *et al.*, *Nature* **584**, 450–456 (2020).
16. L. Dai *et al.*, *Cell* **182**, 722–733.e11 (2020).
17. D. Li *et al.*, *bioRxiv* [Preprint] 2 January 2021. doi:10.1101/2020.12.31.424729.
18. A. R. Shiakolas *et al.*, *Cell Rep. Med.* **2**, 100313 (2021).
19. M. McCallum *et al.*, *Cell* **184**, 2332–2347.e16 (2021).
20. N. Suryadevara *et al.*, *Cell* **184**, 2316–2331.e15 (2021).
21. M. M. Becker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19944–19949 (2008).
22. D. Laczkó *et al.*, *Immunity* **53**, 724–732.e7 (2020).
23. P. Zhou *et al.*, A protective broadly cross-reactive human antibody defines a conserved site of vulnerability on beta-coronavirus spikes. bioRxiv [Preprint] 31 March 2021. doi:10.1101/2021.03.30.437769.
24. K. Lederer *et al.*, *Immunity* **53**, 1281–1295.e5 (2020).
25. D. E. Gordon *et al.*, *Science* **370**, eabe9403 (2020).
26. J. Y. Li *et al.*, *Virus Res.* **286**, 198074 (2020).
27. Y. C. F. Su *et al.*, *mBio* **11**, e01610-20 (2020).
28. B. E. Young *et al.*, *Lancet* **396**, 603–611 (2020).
29. T. P. Sheahan *et al.*, *Nat. Commun.* **11**, 222 (2020).
30. M. E. Schmidt *et al.*, *PLOS Pathog.* **14**, e1006810 (2018).
31. M. Bolles *et al.*, *J. Virol.* **85**, 12201–12215 (2011).
32. L. A. Jackson *et al.*, *N. Engl. J. Med.* **383**, 1920–1931 (2020).
33. E. E. Walsh *et al.*, *N. Engl. J. Med.* **383**, 2439–2450 (2020).
34. L. R. Baden *et al.*, *N. Engl. J. Med.* **384**, 403–416 (2021).
35. K. Wu *et al.*, *N. Engl. J. Med.* **384**, 1468–1470 (2021).
36. M. M. Sauer *et al.*, *Nat. Struct. Mol. Biol.* **28**, 478–486 (2021).
37. A. A. Cohen *et al.*, *Science* **371**, 735–741 (2021).
38. K. O. Saunders *et al.*, *Nature* (2021).
39. L. R. Banner, J. G. Keck, M. M. Lai, *Virology* **175**, 548–555 (1990).
40. J. G. Keck, L. H. Soe, S. Makino, S. A. Stohlman, M. M. Lai, *J. Virol.* **62**, 1989–1998 (1988).

ACKNOWLEDGMENTS

Funding: D.R.M. is currently supported by a Burroughs Wellcome Fund Postdoctoral Enrichment Program Award and a Hanna H. Gray Fellowship from the Howard Hugues Medical Institute and was supported by an NIH NIAID T32 AI007151 and an NIAID F32 AI152296. This research was also supported by funding from the Chan Zuckerberg Initiative awarded to R.S.B. This project was supported by the North Carolina Policy Collaboratory at the University of North Carolina at Chapel Hill, with funding from the North Carolina Coronavirus Relief Fund established and appropriated by the North Carolina General Assembly. This project was funded in part by the National Institute of Allergy and Infectious Diseases (NIAID), NIH, US Department of Health and Human Services awards U01 AI149644, U54 CA260543, AI157155, and AI110700 to R.S.B.; AI124429 and a BioNTech SRA to D.W. and E.N.A.-V.; as well as an animal models contract from the NIH (HHSN272201700036I). Animal histopathology services were performed by the Animal Histopathology and Laboratory Medicine Core at the University of North Carolina, which is supported in part by an NCI Center Core Support Grant (5P30CA016086-41) to the UNC Lineberger Comprehensive Cancer Center. We thank B. L. Mui and Y. K. Tam from Acuitas Therapeutics, Vancouver, BC V6T 1Z3, Canada, for supplying the LNPs. **Author contributions**: D.R.M. and R.S.B. conceived the study. D.R.M. and R.S.B. designed experiments. D.R.M., A.S., S.R.L., and A.W. performed laboratory experiments. N.P. and K.O.S. provided critical reagents. D.R.M., A.S., S.R.L., G.D.I.C., A.W., E.N.A.-V., L.C.L., N.P., R.P., M.B., D.L., B.Y., K.O.S., D.W., B.F.H., and S.A.M. analyzed data and provided critical insight. D.R.M. wrote the first draft of the paper. D.R.M., A.S., S.R.L., G.D.I.C., A.W., E.N.A.-V., L.C.L., N.P., N.P., R.P., M.B., D.L., B.Y., K.O.S., D.W., B.F.H., S.A.M., and R.S.B. read and edited the paper. Funding acquisition: D.R.M. and R.S.B. All authors reviewed and approved the manuscript. **Competing interests:** The University of North Carolina at Chapel Hill has filed provisional patents for which D.R.M. and R.S.B. are co-inventors (US provisional application no. 63/106,247 filed on 27 October 2020) for the chimeric vaccine constructs and their applications described in this study. **Data and materials availability:** The amino acid sequences of the chimeric spike constructs are included in table S1. mRNA sequences are deposited in GenBank with the following accession nos: chimera 1, MZ393687; chimera 2, MZ393688; chimera 3, MZ393689; and chimera 4, MZ393690. Materials generated as part of this study are available from R.S.B. upon request. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/991/suppl/DC1
Materials and Methods
Figs. S1 to S7
Table S1
References (*41*–*43*)
MDAR Reproducibility Checklist

11 March 2021; accepted 15 June 2021
10.1126/science.abi4506

QUALITY CONTROL

Identification of a quality-control factor that monitors failures during proteasome assembly

Eszter Zavodszky[1], Sew-Yeu Peak-Chew[1], Szymon Juszkiewicz[1], Ana J. Narvaez[2]†, Ramanujan S. Hegde[1]*

In eukaryotic cells, half of all proteins function as subunits within multiprotein complexes. Imbalanced synthesis of subunits leads to unassembled intermediates that must be degraded to minimize cellular toxicity. Here, we found that excess PSMC5, a subunit of the proteasome base, was targeted for degradation by the HERC1 ubiquitin ligase in mammalian cells. HERC1 identified unassembled PSMC5 by its cognate assembly chaperone PAAF1. Because PAAF1 only dissociates after assembly, HERC1 could also engage later assembly intermediates such as the PSMC4-PSMC5-PAAF1 complex. A missense mutant of HERC1 that causes neurodegeneration in mice was impaired in the recognition and ubiquitination of the PSMC5-PAAF1 complex. Thus, proteasome assembly factors can serve as adaptors for ubiquitin ligases to facilitate elimination of unassembled intermediates and maintain protein homeostasis.

Many of the cell's multisubunit complexes, such as ribosomes and proteasomes, are exceptionally abundant and contain a large number of subunits. Their maturation can involve dozens or more assembly factors that participate in a multistep pathway (*1*–*3*). The different subunits cannot be produced at a precise desired stoichiometry owing to the inherent noisiness of transcription and translation (*4*, *5*). For highly abundant complexes, even subtle imbalances in subunit synthesis (*6*) can produce an appreciable number of assembly intermediates awaiting the next component. Furthermore, cellular stress and various disease states, most notably cancer, can exaggerate subunit imbalances because of altered or dysregulated gene expression (*6*–*8*). As exemplified by the thalassemias (*9*) or aneuploidy (*10*–*13*), excessive or chronic subunit imbalance can have detrimental gain-of-function consequences. How cells detect stalled orphaned assembly intermediates for selective elimination is not well understood.

[1]MRC Laboratory of Molecular Biology, Cambridge, UK. [2]Discovery Sciences, AstraZeneca, Cambridge, UK.
***Corresponding author. Email: rhegde@mrc-lmb.cam.ac.uk**
†Present address: Moderna Therapeutics, Cambridge, MA, USA.

Identification of degraded candidate orphan proteins

To find prominent orphan proteins caused by imbalanced gene expression, we identified rapidly degraded proteins in the aneuploid breast cancer cell line MCF7. We pulse-labeled nascent proteins for 1 hour with the methionine analog azido-homoalanine (AHA) followed by a chase with methionine for up to 8 hours. The

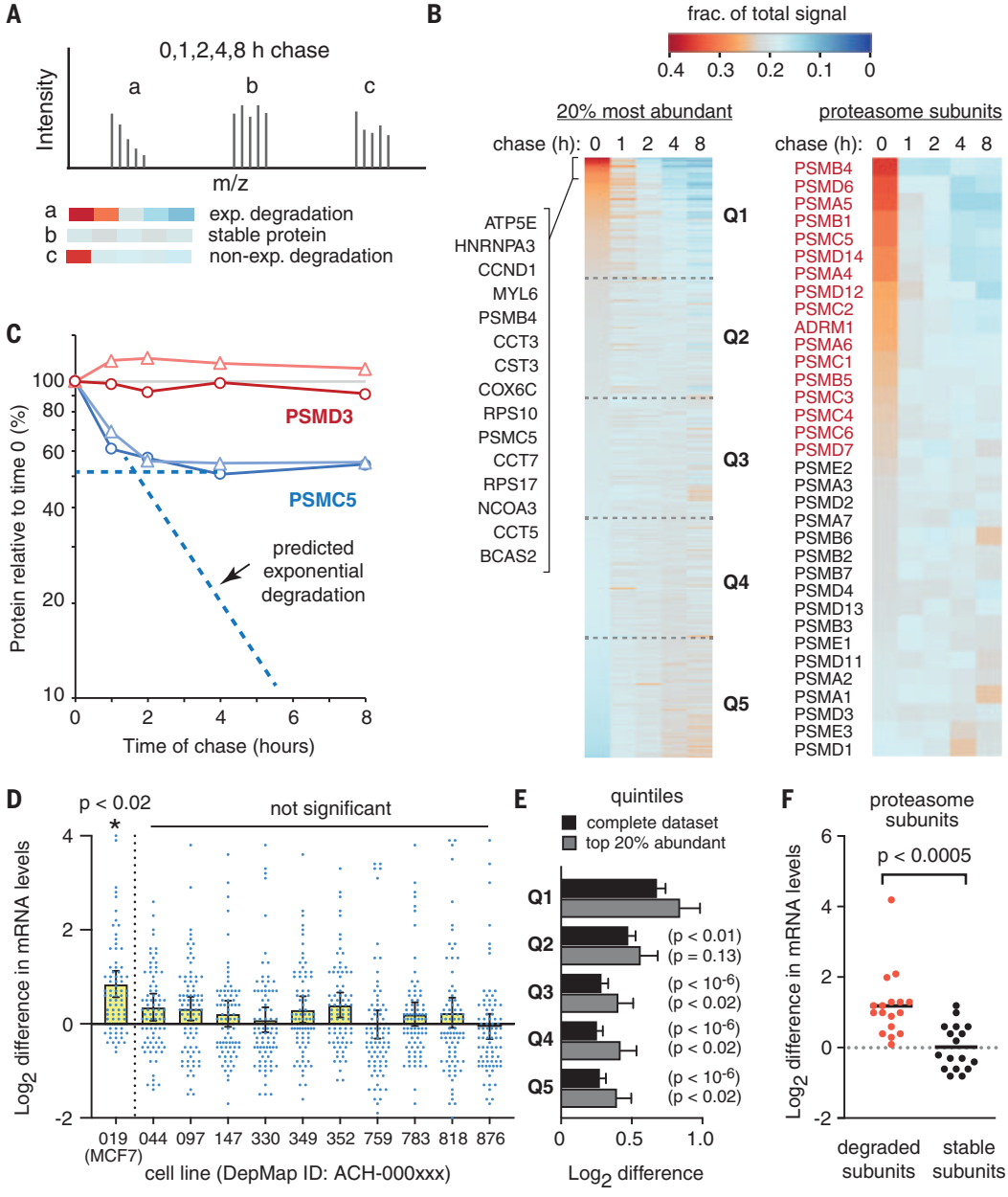


Fig. 1. Identification of rapidly degraded candidate orphans. (**A**) Schematic of how tandem mass tag mass spectrometry data from a pulse chase is displayed as a heat map to visualize a protein's change over time. The relative proportion of the total signal at each time point is displayed in the heat map (red is high; blue is low) for each hypothetical peptide's spectra. exp., exponential. (**B**) Heat maps as in (A) are shown for the 20% most abundant proteins (left) and for subunits of the proteasome (right). Identities of the first 15 proteins of the left heat map are indicated. Each quintile is indicated (Q1 to Q5). (**C**) Data from (B) showing examples of nonexponential degradation (PSMC5) and a stable protein (PSMD3). Each of two replicates is shown in different shades. (**D**) Individual mRNA levels for quintile-1 proteins from (B) (together with their means ± 95% confidence intervals) are shown for 11 cell lines derived from HER2-negative and estrogen receptor–positive breast cancers. Each mRNA level is expressed as the $\log_2$ difference from that mRNA's average level across all 1379 cell lines in the DepMap database. Only MCF7 cells showed a significant ($p < 0.02$) increase of this set of mRNAs, which was also significantly different ($p < 0.02$) from the levels in all the other breast cancer cell lines. (**E**) Plot as in (D) comparing the mRNA levels (means ± SEM) for Q1 through Q5 in MCF7 cells. p values for a two-tailed Student's t test comparison to Q1 are indicated. (**F**) Plot as in (D) for the degraded and stable subunits of the proteasome shown in (B), using the same color-coding as in (B). The black horizontal lines indicate the mean.

AHA-containing proteins across all time points were captured by click chemistry (*14*) and analyzed by quantitative proteomics (Fig. 1A, fig. S1, and table S1). Focusing on the cell's ~450 most highly translated proteins (from 2262 total identified proteins), we found that a minor subset of nascent proteins decreased abruptly in the first hour (Fig. 1B). Notably, 14 of the top 15 such proteins were subunits of multiprotein complexes.

Only a subset of subunits that comprise the proteasome, chaperonin, and ribosome displayed nonexponential degradation, where a proportion of the protein was degraded rapidly and the remainder was comparatively stable (Fig. 1, B and C, and fig. S2). By contrast, degradation of proteins that were not stable subunits of complexes, or the loss of proteins because of secretion, followed exponential decay (fig. S2C). These findings could not be explained by AHA-induced protein misfolding (tables S2 and S3) (*15*) or imbalanced subunit production caused by methionine starvation during AHA labeling (table S4) (*16*).

Instead, the collection of mRNAs for degraded proteins was more highly increased in MCF7 cells than in any of 10 other similar breast cancer cell lines (Fig. 1D) (*17*). The degree of increase for these mRNAs was also greater than that for the set of mRNAs coding for the stable proteins (Fig. 1E). The mRNAs for degraded subunits of the proteasome and ribosome were overrepresented by around 1.5- to 2-fold in MCF7 cells, but not other breast cancer cell lines, relative to the mRNAs for the stable subunits of the respective complex (Fig. 1F and fig. S3). This degree of excess mRNA roughly matched the ~30 to 50% of their encoded proteins that were rapidly degraded. Thus, subunits of multiprotein complexes are major targets for quality control in cancer cells because a population of these proteins evidently become orphaned when they are expressed inappropriately at higher levels than their assembly partners.

HERC1 interacts with and ubiquitinates nascent PSMC5

To identify the quality-control pathway(s) that facilitate orphan degradation, we analyzed

individual subunits of four diverse multiprotein complexes identified as orphans in our proteomics analysis of MCF7 cells. Each epitope-tagged subunit was synthesized in a reticulocyte-based cell-free translation extract with the aim of finding at least one whose recognition for quality control was recapitulated in this system. Two subunits, PSMC5 of the proteasome and CCT3 of the cytosolic chaperonin, were ubiquitinated (Fig. 2A). Both complexes are abundant in reticulocytes (*18*, *19*), explaining why factors needed for their quality

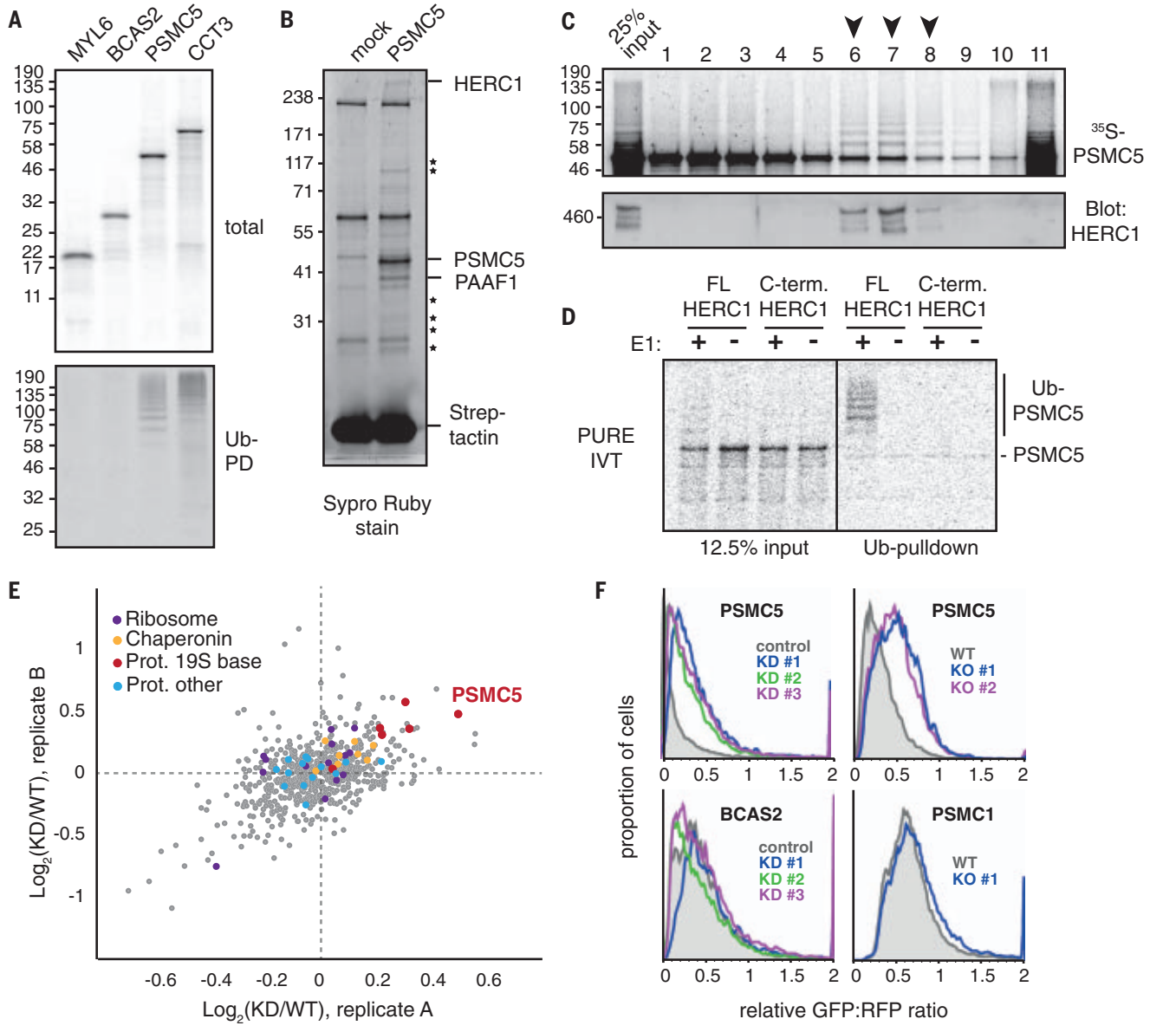


Fig. 2. HERC1 mediates nascent PSMC5 degradation. (**A**) Subunits of four different multimeric complexes were translated in rabbit reticulocyte lysate (RRL) containing ^{35}S-methionine and His-tagged ubiquitin (Ub). PD, pull down. Ubiquitinated translation products were isolated by Ni-nitrilotriacetic acid (Ni-NTA) pull down under denaturing conditions. (**B**) PSMC5 containing a twin-Strep tag (TST) was translated in RRL and affinity-purified under native conditions using immobilized Streptactin, and the products were visualized by SYPRO Ruby stain. Mock indicates a parallel reaction lacking mRNA. The indicated PSMC5-specific interaction partners were identified by mass spectrometry, with stars denoting other 19*S* subunits of the proteasome. (**C**) ^{35}S-methionine–labeled PSMC5 translated in in RRL was separated on a 5 to 25% sucrose gradient. Fractions were analyzed by autoradiography for PSMC5 and by immunoblot for HERC1. Arrows indicate fractions where ubiquitinated PSMC5 cofractionates with HERC1. On this gradient, ~300-kD protein typically migrates in fractions three to five, and a ~500-kD protein in fractions five to seven. (**D**) ^{35}S-methionine–labeled PSMC5 in complex with PAAF1 was produced by translation in the PURE system (see fig. S5A). The complex was mixed with E1, E2, His-Ub, adenosine triphosphate (ATP), and either full-length (FL) or truncated recombinant HERC1 (see fig. S5B). Ubiquitinated products were isolated by Ni-NTA pull down under denaturing conditions and visualized by autoradiography. IVT, in vitro translation. (**E**) MCF7 cells were pretreated with nontargeting (WT) or HERC1-targeting (KD) small interfering RNAs (siRNAs) and metabolically labeled for 1 hour with AHA. The labeled proteins were selectively recovered using click chemistry, and the recovered proteins were analyzed by quantitative mass spectrometry. The top 20% most abundant proteins from two biological replicates are plotted by the KD/WT ratio, with subunits of the proteasome, ribosome, and chaperonin highlighted. (**F**) Flow cytometry analysis of GFP-tagged reporter proteins compared with an internal RFP control. HERC1 was knocked down by three separate siRNAs (KD) or knocked out using CRISPR (KO).

control are present in this system. Conversely, subunits of the spliceosome and myosin, which do not have an appreciable role in reticulocyte biology, were poorly recognized.

Focusing on newly made PSMC5, we observed that its immunoprecipitation under nondenaturing conditions coprecipitated ubiquitin ligase activity (fig. S4A). Large-scale affinity purification of nascent PSMC5 followed by mass spectrometry identified the HECT-domain protein HERC1 as the sole E3 ubiquitin ligase (Fig. 2B and table S5). This interaction was verified by coimmunoprecipitation and immunoblotting (fig. S4B). Furthermore, HERC1 co-fractionated selectively with the ubiquitinated subpopulation of PSMC5 when separated by size by sedimentation through a sucrose gradient (Fig. 2C).

The other interaction partners of nascent PSMC5 suggested that it was engaged in a heterogeneous set of assembly intermediates consistent with PSMC5's broad distribution across the sucrose gradient. The most abundant interaction partner was PSMC5's dedicated chaperone PAAF1 (fig. S4B) (*20*–*23*). Other subunits of the 19*S* base were copurified at lower levels, consistent with partial assembly of PSMC5 with endogenous base subunits present in the lysate (table S5). Assembly beyond this step was evidently less efficient, with incomplete recovery of 19*S* lid subunits and essentially no recovery of 20*S* core subunits.

Engagement of PSMC5 with the 20*S* core would be impaired by the C-terminal epitope tag, thereby ensuring that nearly all nascent PSMC5 would stall at earlier assembly intermediates. Based on the staining intensity of the products (Fig. 2B) and sucrose gradient distribution (Fig. 2C), the most abundant complex was the PSMC5-PAAF1 assembly intermediate. Thus, this unassembled product might be at least one quality-control target of HERC1.

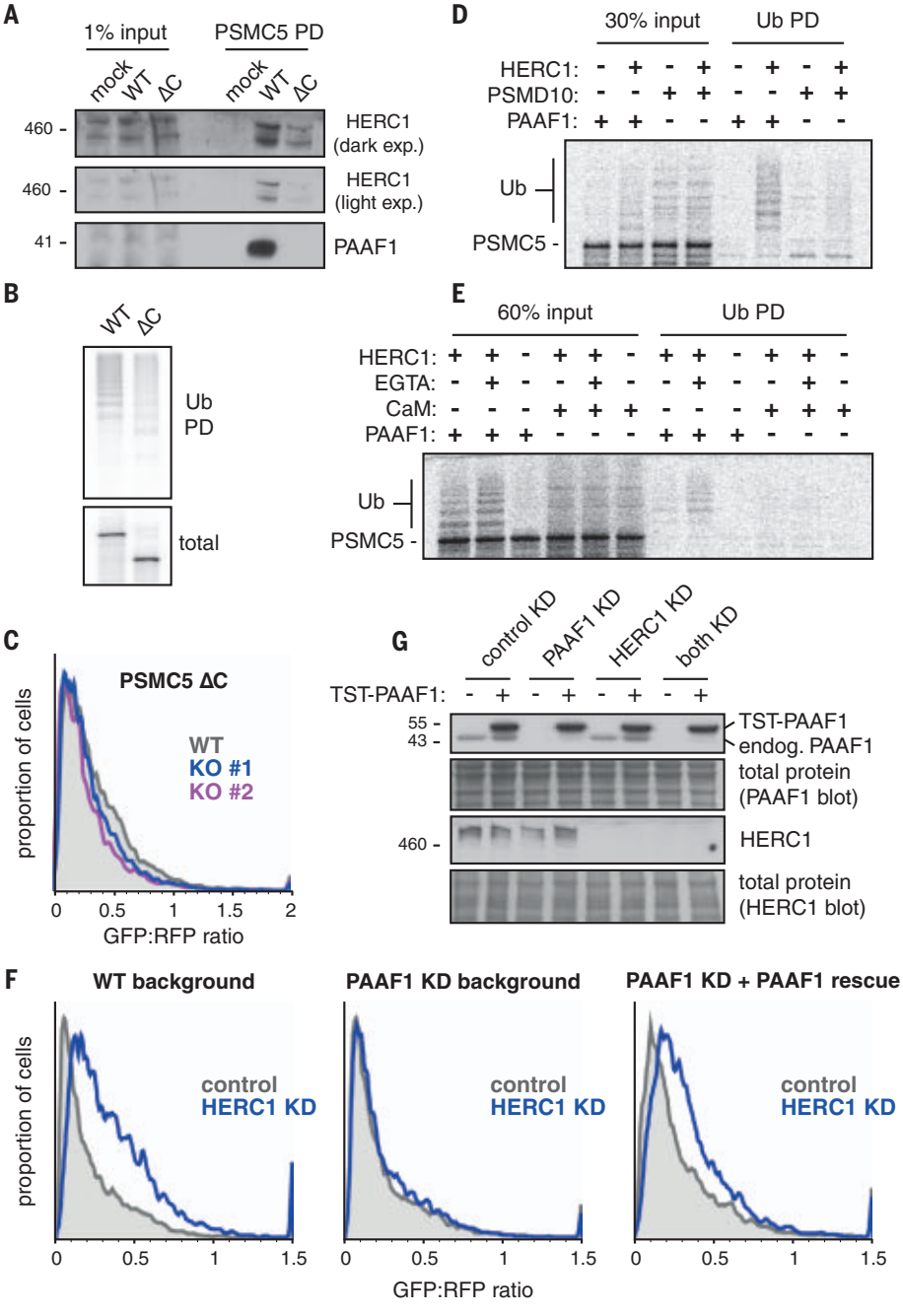


Fig. 3. The assembly chaperone PAAF1 participates in HERC1 target selection. (**A**) Full-length PSMC5 (WT) and PSMC5 lacking its C-terminal 86 residues (ΔC) were translated in RRL and affinity-purified under native conditions using a C-terminal TST. The input and purified samples were blotted for HERC1 and PAAF1. SYPRO-Ruby staining (fig. S10B) was used to verify equal recovery of the translation products. (**B**) Full-length and truncated PSMC5 were translated in RRL in the presence of ^{35}S-methionine and His-tagged ubiquitin (Ub). Ubiquitinated products were isolated by Ni-NTA pull down. (**C**) Flow cytometry analysis of GFP-tagged C-terminally truncated PSMC5 in WT and HERC1 KO MCF7 cells. (**D**) ^{35}S-methionine–labeled PSMC5 in complex with PAAF1 or PSMD10 was prepared by translation in the PURE system. The complexes were mixed with E1, E2, His-Ub, ATP, and full-length recombinant HERC1. Ubiquitinated products were isolated by Ni-NTA pull down under denaturing conditions. (**E**) ^{35}S-methionine–labeled PSMC5 in complex with PAAF1 or calmodulin (CaM) was prepared and mixed with ubiquitination components, as in (D). Where indicated, EGTA was used to dissociate CaM from PSMC5. Ubiquitinated products were isolated by Ni-NTA pull down under denaturing conditions. (**F**) Flow cytometry analysis of GFP-tagged PSMC5 in MCF7 cells. HERC1 and PAAF1 were knocked down separately or simultaneously by single siRNA oligonucleotides. TST-PAAF1 with siRNA-resistant silent mutations was reintroduced where indicated (PAAF1 rescue). (**G**) KD of endogenous PAAF1 and HERC1, as well as rescue with TST-PAAF1, was verified by immunoblot for the experiment in (F).

PSMC5 synthesized in a translation system reconstituted from purified recombinant *Escherichia coli* translation factors (known as the PURE system) was entirely insoluble unless a chaperone, such as recombinant PAAF1, was included during translation (fig. S5A). The PSMC5-PAAF1 heterodimer from the peak fractions of the gradient could be ubiquitinated by full-length recombinant HERC1 at physiologic concentrations, but not by a C-terminal domain that retains full E3 ligase activity (Fig. 2D and fig. S5B). Thus, HERC1 can recognize the unassembled PSMC5-PAAF1 complex and ubiquitinate PSMC5.

HERC1 facilitates degradation of PSMC5 in cells

Pulse-chase experiments using ^{35}S-methionine showed that roughly half of newly synthesized PSMC5 was degraded shortly after synthesis in MCF7 cells (fig. S6A). PSMC5 degradation was blunted in cells knocked out for HERC1 (fig. S6B). Proteomic analysis of nascent proteins synthesized during 1 hour in untreated versus HERC1 knockdown (KD) cells revealed that PSMC5 was overrepresented (~1.4-fold) in KD cells (Fig. 2E and table S6). Relatively few other proteins were overrepresented similarly, indicating that HERC1 does not generally influence protein synthesis or degradation. For example, chaperonin subunits, ribosomal proteins, and most subunits of the proteasome were unaffected (Fig. 2E). By contrast, five of six adenosine triphosphatase (ATPase) subunits (including PSMC5) of the 19*S* proteasome base were overrepresented in the KD sample, suggesting that this class of proteins was selectively influenced by HERC1.

Although the burden of surplus PSMC5 and other subunits is exaggerated in MCF7 cells, ~10 to 20% imbalanced production can occur in normal cells owing to gene expression noise (*4*, *5*). Indeed, 19*S* ATPase subunits were modestly (by ~21%) but significantly stabilized upon HERC1 KD in the noncancer breast epithelial cell line MCF10a (fig. S7 and table S7). Subunits of the cytosolic chaperonin (CCTs), which are not targets for HERC1, were not changed in HERC1 KD cells. Thus, the problem of subunit imbalance, although enhanced by genomic dysregulation in cancer cells, is nonetheless appreciable in noncancer cells.

These proteomic observations were examined in focused assays using fluorescent protein reporters. Reporter translation produces two fluorescent proteins: green fluorescent protein (GFP) that is fused to a protein of interest and red fluorescent protein (RFP). The GFP:RFP ratio provides a quantitative assessment of any changes to the GFP-tagged protein's stability (*24*). Overexpressed N-terminally–tagged PSMC5 was mostly degraded but was stabilized by KD or knockout (KO) of HERC1 in MCF7 cells (Fig. 2F and fig. S8).

Furthermore, degradation of excess subunits of other complexes or a subunit of the 20*S* proteasome core was not affected by HERC1 KD (Fig. 2F and fig. S9A). Surprisingly, other 19*S* ATPase ring members that were affected by HERC1 KD in the proteomics experiment were neither degraded as effectively as PSMC5 nor affected strongly by HERC1 KO when tested by exogenous overexpression in the reporter assay (Fig. 2F and fig. S9B). This suggested that the effect of HERC1 on non-PSMC5 base subunits could be indirect, an idea consistent with relatively poor interaction of most of these subunits with HERC1 in vitro despite their considerable sequence and structural homology to each other (fig. S10A).

PAAF1 is required for recognition of PSMC5 by HERC1

To understand why overexpressed PSMC5, but not other base subunits, was degraded via HERC1, we investigated the mechanism of HERC1 target selection. Although each subunit of the 19*S* ATPase ring is similar, their C-terminal domains engage different assembly factors (*20*–*23*, *25*). Recognizing that our reconstituted ubiquitination experiment (Fig. 2D) used the PSMC5-PAAF1 complex as a substrate (fig. S5), we tested whether PAAF1 plays a role in HERC1 recognition. Indeed, in vitro–translated PSMC5 lacking the C-terminal domain (PSMC5ΔC) did not bind PAAF1, showed reduced HERC1 binding, and showed reduced ubiquitination compared with full-length PSMC5 (Fig. 3, A and B).

Although overexpressed PSMC5ΔC was degraded in MCF7 cells, its degradation did not depend on HERC1 (Fig. 3C). This suggested that HERC1 either recognizes the C terminus, recognizes PAAF1, or that non-native interaction partners of PSMC5ΔC (fig. S10B) obscure HERC1 access. To distinguish between these possibilities, we tested PSMC5 ubiquitination when it is complexed with another 19*S* assembly chaperone (PSMD10), is complexed with the nonphysiologic chaperone-like protein calmodulin (*26*), or is uncomplexed with any factor. PSMC10 ordinarily chaperones the C-terminal domain of PSMC4 but can engage PSMC5 in the purified translation reaction owing to the homology between their C-terminal domains. Calmodulin, by contrast, has broad Ca^{2+}-dependent binding activity. Both proteins precluded nascent PSMC5 aggregation sufficiently well in the PURE system to provide substrate complexes for HERC1 ubiquitination assays.

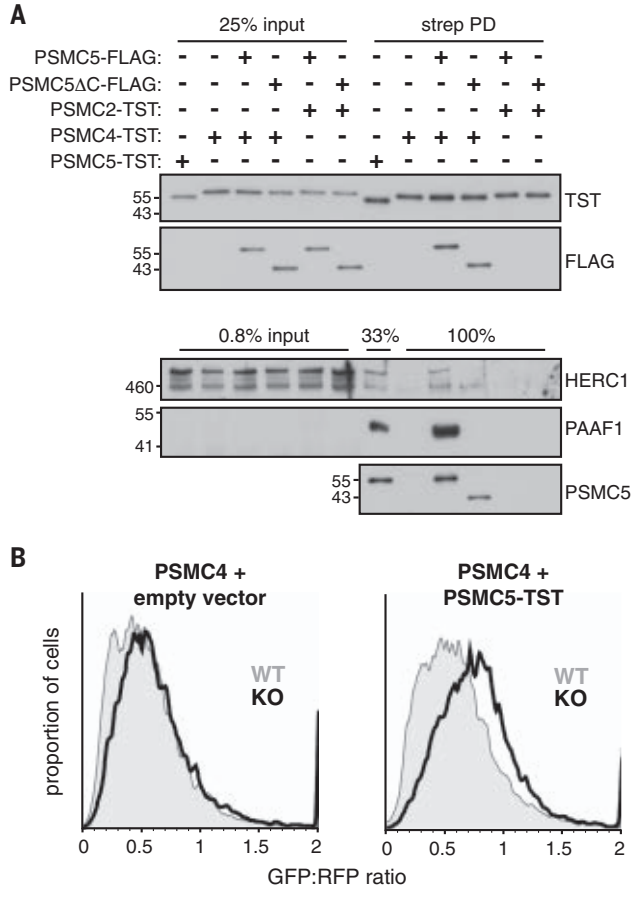


Fig. 4. HERC1 recognizes multiple assembly intermediates using PAAF1. (**A**) FLAG-tagged full-length or C-terminally truncated (ΔC) PSMC5 was cotranslated with a twin-Strep-tagged partner ATPase (PSMC4) or distal ring member (PSMC2) in RRL. Complexes were affinity purified using the TST and blotted for TST, FLAG, HERC1, PAAF1, and PSMC5 as indicated. TST-PSMC5 was translated and purified directly as a positive control. Reactions containing only one ring member included mRNA for an irrelevant protein (the cytosolic domain of Sec61β) to ensure equal rates of synthesis. The relative amounts of the different samples loaded for the two sets of blots are indicated. (**B**) Flow cytometry analysis of GFP-tagged PSMC4 in WT and HERC1 KO MCF7 cells, with and without cotransfection of nonfluorescent PSMC5.

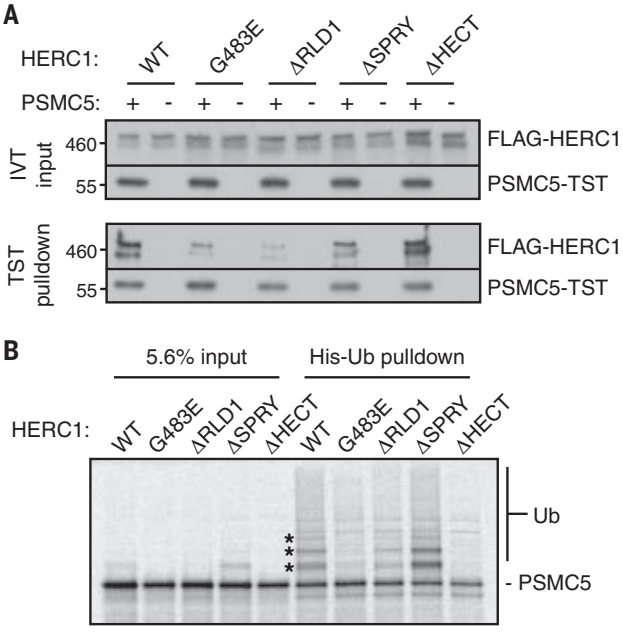


Fig. 5. Disease-causing HERC1 mutant is deficient in PSMC5 recognition and ubiquitination. (**A**) Recombinant purified FLAG-tagged HERC1 proteins (see fig. S14) were added at 2 nM to RRL in vitro translation reactions with or without mRNA for twin-Strep-tagged PSMC5. PSMC5-TST was affinity purified under native conditions using Strep-tactin, and inputs and elutions were visualized by immunoblotting. (**B**) ^{35}S-methionine–labeled PSMC5 in complex with PAAF1 was produced by translation in the PURE system (as in fig. S5A). The complex was mixed with E1, E2, His-Ub, ATP, and 8 nM of WT or mutant HERC1. Ubiquitinated products were enriched by Ni-NTA pull down under denaturing conditions. The input and ubiquitin pull-down samples were visualized by autoradiography. The positions of mono-, di-, and tri-ubiquitinated PSMC5 are indicated with asterisks (also visible faintly in the input sample).

HERC1 did not effectively ubiquitinate either the PSMC5-PSMD10 complex or the PSMC5-calmodulin complex (Fig. 3, D and E). EGTA-mediated release of calmodulin to fully expose PSMC5 to HERC1 also did not permit ubiquitination. Thus, neither uncomplexed PSMC5 nor PSMC5 complexed with PSMD10 or calmodulin was recognized by HERC1, illustrating a crucial role for PAAF1 in recognition. Indeed, HERC1 depletion failed to stabilize PSMC5 in cells depleted of PAAF1 (Fig. 3F). Efficient PSMC5 recognition by HERC1 therefore relies on PSMC5's chaperone PAAF1. PSMC5 produced in the absence of PAAF1 is presumably misfolded (consistent with its aggregation in the PURE system), triggering PSMC5 degradation by a HERC1-independent pathway. This further indicates that HERC1 is not simply recognizing PSMC5 misfolding but rather is recognizing a PAAF1-containing putative assembly intermediate.

HERC1 interacts with a PSMC4-containing assembly intermediate

PAAF1 remains bound to the C-terminal domain of PSMC5 throughout its assembly with other ATPase subunits of the 19*S* base, dissociating concomitant with C-terminal insertion into the 20*S* core particle (*21*, *23*). Hence, assembly intermediates downstream of the initial PSMC5-PAAF1 complex are potential targets for HERC1, possibly explaining why other base subunits are affected by HERC1 KD in the context of PSMC5 excess. To test this idea, we determined if the PSMC5-PSMC4 assembly intermediate is targeted by HERC1 using retained PAAF1.

In vitro–translated PSMC4 interacted equally well with coexpressed PSMC5 and PSMC5ΔC (Fig. 4A, top), verifying that each nascent subunit is assembly-competent. Affinity purification of PSMC4 coprecipitated PAAF1 and HERC1 only from PSMC5-containing reactions, but not from reactions containing PSMC5ΔC or lacking PSMC5 (Fig. 4A, bottom). Thus, PSMC4 associates with HERC1 through the PSMC5-PAAF1 complex. In accordance with HERC1 recruitment, the PSMC5-PSMC4 intermediate remains vulnerable to ubiquitination (fig. S11). Thus, assembly of PSMC5 with PSMC4 does not protect PSMC5 from ubiquitination, presumably because retained PAAF1 is still able to recruit HERC1 to the complex.

Although overexpressed fluorescent PSMC4 is mostly degraded independently of HERC1, coexpression with nonfluorescent PSMC5 imparts partial HERC1 dependence to PSMC4 degradation (Fig. 4B). This provides in vivo support for HERC1 targeting PSMC4 by recognition of the associated PSMC5-PAAF1 complex. These observations provide an explanation for partial PSMC4 stabilization in MCF7 cells knocked down for HERC1 (Fig. 2E). Although we have not examined other assembly intermediates, it is plausible that later PAAF1-containing complexes are also potential targets for HERC1.

Why does HERC1 not interfere with normal assembly? One model is kinetic competition (*27*, *28*), with assembly occurring faster than HERC1 recognition. This mechanism is plausible because HERC1 is ~10-fold less abundant than PAAF1 (*29*, *30*). Boosting HERC1 levels using CRISPR activation (*31*) of the endogenous *HERC1* promoter led to promiscuous degradation of endogenous nascent PSMC5 and PSMC4 (fig. S12), supporting a competition-based mechanism.

Disease-causing HERC1 mutant is deficient in PSMC5 recognition

Coimmunoprecipitation experiments using a series of domain deletion constructs expressed in cultured cells showed that the first RCC1-like domain (RLD1) in HERC1 is important for interaction with PSMC5 (fig. S13). This finding was illuminating because the recessive neurodegeneration-causing *tambaleante* allele in mouse was mapped to a Gly^{483}→Glu (G483E) missense mutation in RLD1 of HERC1. Biochemical analysis of purified HERC1 variants (fig. S14) showed that the ΔRLD1 deletion mutant and G483E point mutant were impaired in their interaction with nascent PSMC5, whereas other deletion mutants were not (Fig. 5A and fig. S15A).

Ubiquitination assays using the PSMC5-PAAF1 complex produced in the PURE system showed a marked reduction of PSMC5 ubiquitination by the ΔRLD1 and G483E mutants (Fig. 5B and fig. S15B) despite no impairment of autoubiquitination (fig. S15C). This suggests that G483E is specifically deficient in interaction with and ubiquitination of PSMC5. This mutant is expressed normally in mouse (*32*) and cultured cells with comparable biochemical behavior as the wild-type (WT) protein (fig. S15, D and E), illustrating that the mutant protein is not grossly misfolded or eliminated by cellular quality control. These results suggest that a deficiency in HERC1-mediated quality control contributes to the disease phenotype of *tambaleante* mice. Consistent with this idea, the impaired cell types in mice are those that most rely on effective proteostasis (*33*), especially Purkinje cells, which are frequently lost with perturbed proteostasis (*34*–*36*).

Conclusions and perspective

Unassembled subunits have long been appreciated to be quality-control targets for one of two reasons: either they misfold in the absence of assembly or the exposed assembly interface is recognized by quality-control factors (*37*–*40*). In both cases, exposed hydrophobic surfaces are thought to be the target for quality control. Our findings reveal a qualitatively different mechanism: Delayed assembly cues ubiquitin ligase recruitment using the associated assembly factor. A major advantage for the cell of such a kinetic competition mechanism is that recognition does not depend on potentially toxic misfolding or aggregation. Instead, a simple delay would be sufficient to trigger elimination of otherwise normal intermediates. By using an assembly factor to identify incomplete

products, a ubiquitin ligase can potentially recognize multiple intermediates along the assembly pathway.

HERC1 is both widely expressed and widely conserved (*41*, *42*). Loss of HERC1 function in mice and humans leads to neurological defects (*32*, *43*–*46*) associated with deficient proteostasis (*33*). Our finding that a neurodegeneration-causing missense mutation is strongly impaired in PSMC5 ubiquitination in a purified system implicates this quality-control pathway in maintaining proteostasis in vivo. More generally, the insights provided here should motivate systematic searches for assembly factor–dependent ubiquitin ligases that eliminate stalled intermediates of other major cellular complexes. Because orphan proteins and intermediates are especially prominent in cancer cells with aberrant gene expression, mechanisms to eliminate these products may be important for their rapid growth. Intriguingly, HERC1 is up-regulated in various cancer cell lines and tumors (*42*), and its overexpression can provide a selective advantage to tumor growth and metastasis in mouse (*47*). RNA sequencing of a wide range of cancer-derived cells (*17*) indicates that different populations of orphans feature in different tumors. Thus, identifying the quality-control pathways for other abundant orphans using the approaches defined here may provide useful therapeutic targets.

REFERENCES AND NOTES

1. K. E. Bohnsack, M. T. Bohnsack, *EMBO J.* **38**, e100278 (2019).
2. S. Klinge, J. L. Woolford Jr., *Nat. Rev. Mol. Cell Biol.* **20**, 116–131 (2019).
3. A. Rousseau, A. Bertolotti, *Nat. Rev. Mol. Cell Biol.* **19**, 697–712 (2018).
4. J. M. Raser, E. K. O'Shea, *Science* **309**, 2010–2013 (2005).
5. B. Munsky, G. Neuert, A. van Oudenaarden, *Science* **336**, 183–187 (2012).
6. J. W. Harper, E. J. Bennett, *Nature* **537**, 328–338 (2016).
7. Y.-C. Tang, A. Amon, *Cell* **152**, 394–405 (2013).
8. D. Hanahan, R. A. Weinberg, *Cell* **144**, 646–674 (2011).
9. R. Origa, *Genet. Med.* **19**, 609–619 (2017).
10. C. M. Brennan *et al.*, *Genes Dev.* **33**, 1031–1047 (2019).
11. N. Dephoure *et al.*, *eLife* **3**, e03023 (2014).
12. E. M. Torres *et al.*, *Science* **317**, 916–924 (2007).
13. B. R. Williams *et al.*, *Science* **322**, 703–709 (2008).
14. D. C. Dieterich, A. J. Link, J. Graumann, D. A. Tirrell, E. M. Schuman, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9482–9487 (2006).
15. E. McShane *et al.*, *Cell* **167**, 803–815.e21 (2016).
16. C. C. Thoreen *et al.*, *Nature* **485**, 109–113 (2012).
17. M. Ghandi *et al.*, *Nature* **569**, 503–508 (2019).
18. N. J. Cowan, *Methods Enzymol.* **290**, 230–241 (1998).
19. J. Driscoll, A. L. Goldberg, *J. Biol. Chem.* **265**, 4789–4792 (1990).
20. M. Funakoshi, R. J. Tomko Jr., H. Kobayashi, M. Hochstrasser, *Cell* **137**, 887–899 (2009).
21. T. Kaneko *et al.*, *Cell* **137**, 914–925 (2009).
22. J. Roelofs *et al.*, *Nature* **459**, 861–865 (2009).
23. Y. Saeki, A. Toh-E, T. Kudo, H. Kawamura, K. Tanaka, *Cell* **137**, 900–913 (2009).
24. E. Itakura *et al.*, *Mol. Cell* **63**, 21–33 (2016).
25. S. Park *et al.*, *Nature* **459**, 866–870 (2009).
26. S. Shao, R. S. Hegde, *Cell* **147**, 1576–1588 (2011).
27. S. Shao, R. S. Hegde, *Trends Biochem. Sci.* **41**, 124–137 (2016).
28. S. Wickner, M. R. Maurizi, S. Gottesman, *Science* **286**, 1888–1893 (1999).
29. D. N. Itzhak, S. Tyanova, J. Cox, G. H. Borner, *eLife* **5**, e16950 (2016).
30. N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, M. Mann, *Nat. Methods* **11**, 319–324 (2014).
31. S. Konermann *et al.*, *Nature* **517**, 583–588 (2015).
32. T. Mashimo *et al.*, *PLOS Genet.* **5**, e1000784 (2009).
33. R. Ruiz, E. M. Pérez-Villegas, S. Bachiller, J. L. Rosa, J. A. Armengol, *Front. Neuroanat.* **10**, 42 (2016).
34. T. Hara *et al.*, *Nature* **441**, 885–889 (2006).
35. L. Zhao, C. Longo-Guess, B. S. Harris, J.-W. Lee, S. L. Ackerman, *Nat. Genet.* **37**, 974–979 (2005).
36. J. W. Lee *et al.*, *Nature* **443**, 50–55 (2006).
37. A. Shemorry, C.-S. Hwang, A. Varshavsky, *Mol. Cell* **50**, 540–551 (2013).
38. M.-K. Sung *et al.*, *eLife* **5**, e19105 (2016).
39. Y. Xu, D. E. Anderson, Y. Ye, *Cell Discov.* **2**, 16040 (2016).
40. K. Yanagitani, S. Juszkiewicz, R. S. Hegde, *Science* **357**, 472–475 (2017).
41. J. García-Cano, A. Martinez-Martinez, J. Sala-Gaston, L. Pedrazza, J. L. Rosa, *Front. Physiol.* **10**, 1014 (2019).
42. J. L. Rosa, R. P. Casaroli-Marano, A. J. Buckler, S. Vilaró, M. Barbacid, *EMBO J.* **15**, 4262–4273 (1996).
43. S. Aggarwal, A. D. Bhowmik, V. L. Ramprasad, S. Murugan, A. Dalal, *Am. J. Med. Genet. A.* **170**, 1868–1873 (2016).
44. L. S. Nguyen *et al.*, *Eur. J. Hum. Genet.* **24**, 455–458 (2016).
45. O. Ortega-Recalde *et al.*, *Clin. Genet.* **88**, e1–e3 (2015).
46. G. E. Utine *et al.*, *Eur. J. Med. Genet.* **60**, 279–283 (2017).
47. O. Fasano, D. Birnbaum, L. Edlund, J. Fogh, M. Wigler, *Mol. Cell. Biol.* **4**, 1695–1705 (1984).

ACKNOWLEDGMENTS

We thank members of the Hegde lab for productive discussions and comments on this manuscript; L. Figueiredo for help with heatmaps; M. Daly and F. Zhang for help with flow cytometry; and M. Kivlen for preparation of the PURE translation system. **Funding:** Funding was provided by UK Medical Research Council grant MC_ UP_A022_1007 (R.S.H.) and LMB-AZ BlueSky project BSF27 (A.J.N. and R.S.H.). **Author contributions:** E.Z. performed nearly all of the experiments. S.-Y.P.-C. performed mass spectrometry and contributed to its analysis. S.J. helped with mapping the substrate-binding domains of HERC1. E.Z., A.J.N., and R.S.H. conceived the project. E.Z. and R.S.H. wrote the manuscript. A.J.N. and R.S.H. provided project supervision. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Reagents used in this study are available from R.S.H. upon request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/998/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S7
References (*48*–*53*)
MDAR Reproducibility Checklist

6 May 2020; resubmitted 22 June 2021
Accepted 15 July 2021
10.1126/science.abc6500

ORGANIC CHEMISTRY

Photomediated ring contraction of saturated heterocycles

Justin Jurczyk[1]†, Michaelyn C. Lux[2]†, Donovon Adpressa[3], Sojung F. Kim[1], Yu-hong Lam[4]*, Charles S. Yeung[2]*, Richmond Sarpong[1]*

Saturated heterocycles are found in numerous therapeutics and bioactive natural products and are abundant in many medicinal and agrochemical compound libraries. To access new chemical space and function, many methods for functionalization on the periphery of these structures have been developed. Comparatively fewer methods are known for restructuring their core framework. Herein, we describe a visible light–mediated ring contraction of α-acylated saturated heterocycles. This unconventional transformation is orthogonal to traditional ring contractions, challenging the paradigm for diversification of heterocycles including piperidine, morpholine, thiane, tetrahydropyran, and tetrahydroisoquinoline derivatives. The success of this Norrish type II variant rests on reactivity differences between photoreactive ketone groups in specific chemical environments. This strategy was applied to late-stage remodeling of pharmaceutical derivatives, peptides, and sugars.

Saturated heterocycles, especially piperidines, are among the most frequently encountered scaffolds in biologically active small molecules (*1*, *2*). The prevalence of the piperidine scaffold in pharmaceutical and agrochemical compound libraries has resulted in increased interest in the development of methods for its site-selective derivatization (Fig. 1A) to access new chemical space. Recent advancements in C(sp^3)–H functionalization have provided many powerful methods for peripheral functionalization of the piperidine framework (*3*–*12*), shifting the paradigm in late-stage functionalization. By comparison, methods that modify the piperidine skeleton are less abundant, prompting a quest for methods for this purpose. Such methods would be especially valuable for drug discovery, in which maximizing structural diversity, specifically at a late stage, is highly valued (*13*). To this end, we sought to develop a ring contraction of piperidines and related saturated heterocycles to carbocyclic frameworks to achieve core, rather than peripheral, modification.

Relatively few methods for piperidine ring contraction exist compared with those available for the ring contraction of carbocycles bearing functional handles. For example, in

[1]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [2]Discovery Chemistry, Merck & Co., Inc., Boston, MA 02115, USA. [3]Analytical Research and Development, Merck & Co. Inc., Boston, MA 02115, USA. [4]Computational and Structural Chemistry, Merck & Co. Inc., Rahway, NJ 07065, USA.
***Corresponding author. Email: yu.hong.lam@merck.com (Y.-h.L.); charles.yeung@merck.com (C.S.Y.); rsarpong@berkeley.edu (R.S.)**
†These authors contributed equally to this work.

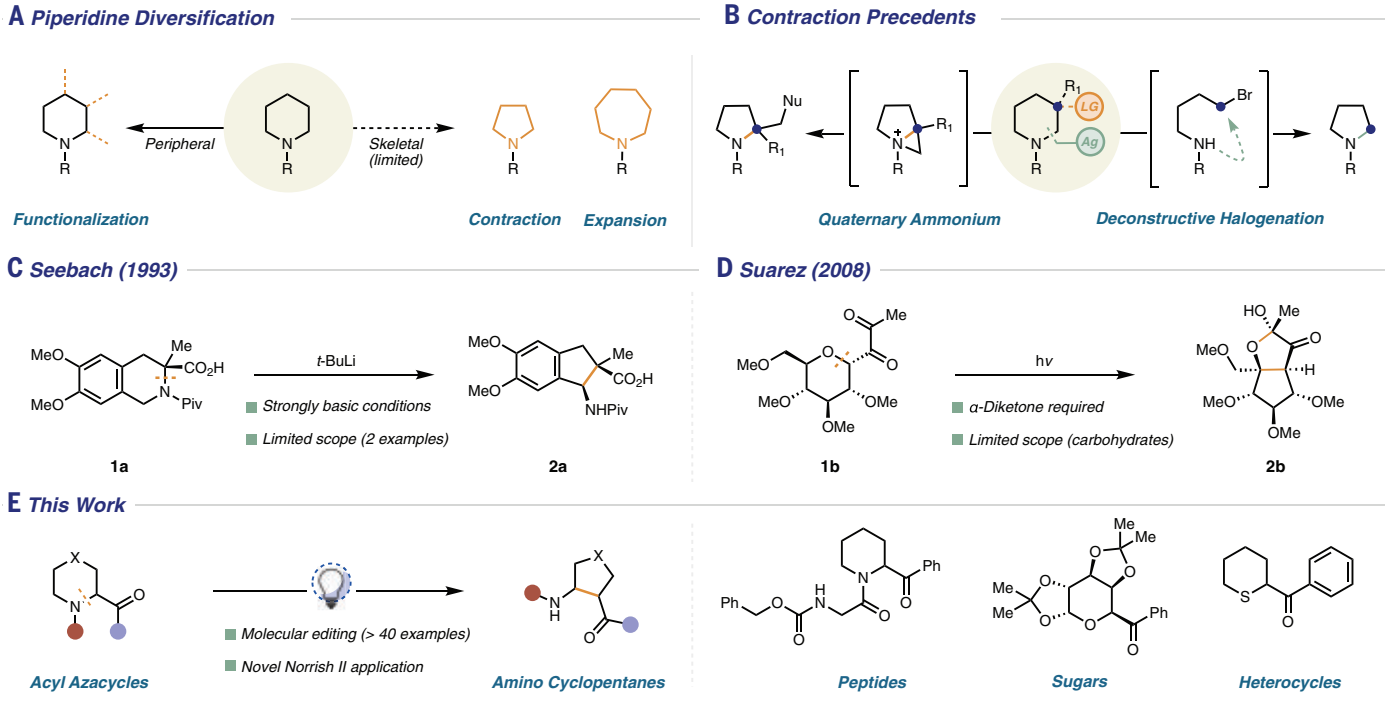


Fig. 1. Approaches to piperidine diversification. (**A**) Peripheral functionalization and skeletal remodeling. (**B**) Selected examples of ring contractions on piperidine frameworks. (**C**) Seminal report of Seebach and co-workers' unusual THIQ ring contraction (*20*). (**D**) Contraction of carbohydrates reported by Suárez and co-workers (*21*). (**E**) Norrish type II approach to piperidine skeletal framework modification (this work).

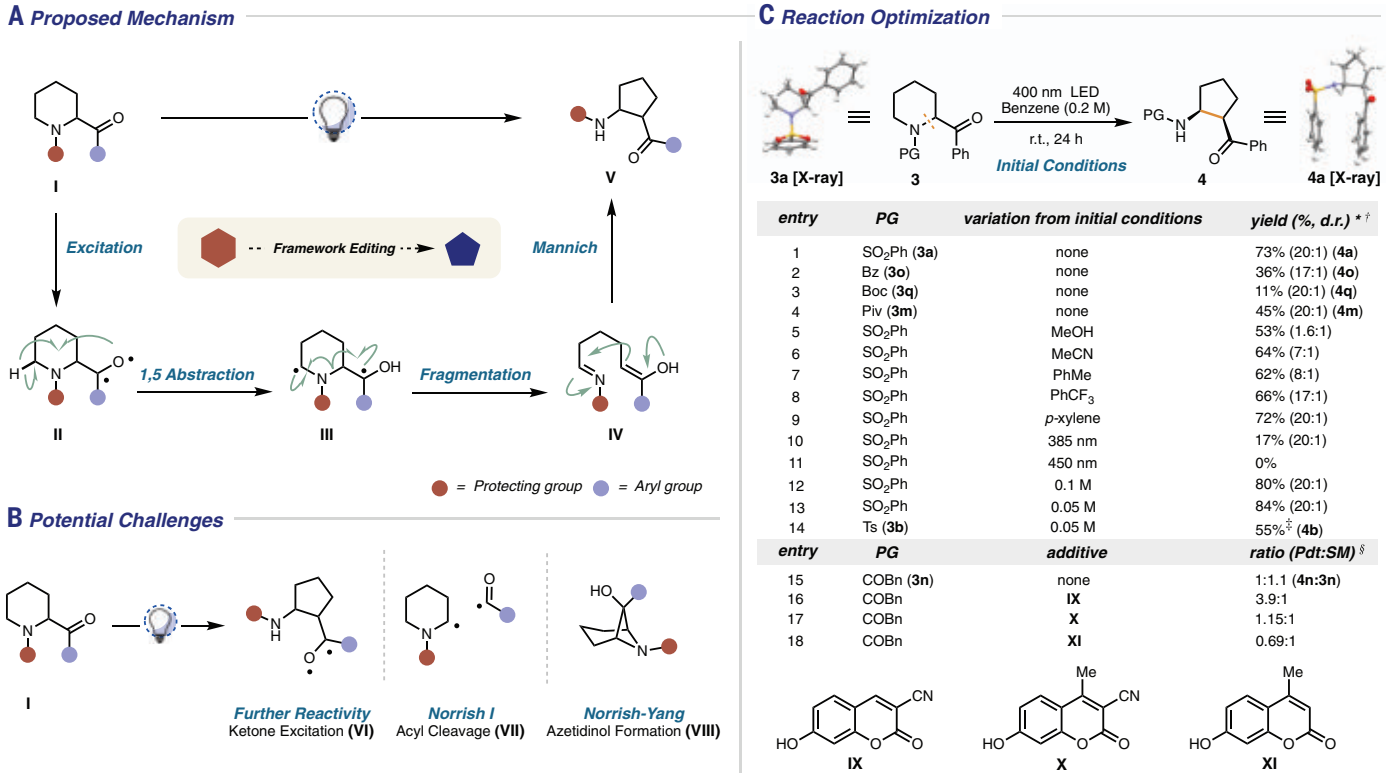
cyclohexanone derivatives alone, anionic, carbene, and cationic intermediates have all been exploited to achieve ring contraction. By contrast, for saturated nitrogen–containing heterocycles (azacycles), commonly used ring contraction strategies predominantly leverage bicyclic, quaternary-ammonium intermediates, which undergo a formal ring contraction upon attack from an exogenous nucleophile (Fig. 1B) (*14*, *15*). More recently, we reported an approach for piperidine ring contraction through oxidative C(sp^3)–N bond cleavage, wherein silver-mediated deconstructive bromination of *N*-benzoyl piperidines followed by intramolecular C–N bond reformation in the resulting acyclic bromoamine furnishes the corresponding *N*-benzoyl pyrrolidine scaffolds in two steps (Fig. 1B) (*16*). Although useful, these tactics are limited to piperidine-to-pyrrolidine scaffold conversions, require specific substitution patterns (*14*, *15*), and use strongly oxidizing conditions that pose a challenge for late-stage diversification (*16*, *17*). The ring contractions of unsaturated azacyclic systems such as pyridiniums and dihydropyridines have also been explored, for example, by using light (*18*, *19*). However, these transformations are limited to substrates with well-recognized photoreactivity profiles.

We sought to develop a complementary and mechanistically distinct approach to piperidine ring contraction that could also be applied to a wide range of saturated heterocycles to provide access to underexplored, skeletally diverse substrates. We drew inspiration from an unusual transformation reported by Seebach and co-workers (Fig. 1C), in which an α-carboxyl tetrahydroisoquinoline (THIQ) derivative (**1a**) underwent contraction to the corresponding indane scaffold (**2a**) under strongly basic conditions. This transformation features a curious endo-to-exocyclic nitrogen atom transposition, giving rise to a β–amino acid structural motif (*20*). Despite its novelty, only two examples were reported, and the requisite strongly basic conditions limit its potential application to skeletal modification of drug-like molecules bearing base-reactive functional groups. A more functional-group-compatible, albeit similarly specific, variant of this type of transformation was reported by Suárez and co-workers for α-diketonyl sugars (e.g., **1b**), which undergo light-mediated ring contraction with a subsequent hemiketalization to give fused [5,5] ring systems (i.e., **2b**) (Fig. 1D) (*21*, *22*). Notably, the α-diketone moiety required for the reaction is derivatized in the resulting product. Our design of a complementary ring contraction sought broad functional group compatibility and a widened scope beyond α-diketonyl-derived sugars. Herein, we report the visible light–mediated ring contraction of α-acylated cyclic piperidines to furnish *cis*-1,2–disubstituted cyclopentane scaffolds (Fig. 1E) and the extension of this method to other saturated heterocycles including tetrahydropyrans and thianes. The success of these Norrish type II transformations hinged on predicted (and observed) photophysical differences between the ketone groups in the starting substrates and products (vide infra). An asymmetric variant of this transformation is also described.

Development of photomediated ring contraction

We commenced our studies by focusing on piperidine ring contractions (Fig. 2A). We envisioned that upon irradiation, α-acylated precursors such as **I**, in which the aroyl group is disposed pseudoaxially to avoid pseudo $A^{1,3}$-like strain (*23*), would undergo excitation and intersystem crossing to afford **II** in the triplet state. A subsequent Norrish type II 1,5-hydrogen atom transfer (HAT) would yield the corresponding 1,4-diradical (**III**), which would undergo homolytic C–N bond fragmentation, leading to imine-enol **IV**. The desired cyclopentane product would then result from an intramolecular enol attack on the tethered imine (Mannich reaction) to afford cyclopentane **V**. Under photoirradiation, however, we also recognized the potential for undesired reactivity (Fig. 2B), especially further reactivity of the reaction product (**V**) to form additional excited species such as **VI**.

In principle, the success of the reaction would depend on subtle differences in reactivity between the starting material (**I**) and the reaction product (**V**; an anticipated "photostationary" state), because both bear a



entry	PG	variation from initial conditions	yield (%, d.r.) *†
1	SO_2Ph (**3a**)	none	73% (20:1) (**4a**)
2	Bz (**3o**)	none	36% (17:1) (**4o**)
3	Boc (**3q**)	none	11% (20:1) (**4q**)
4	Piv (**3m**)	none	45% (20:1) (**4m**)
5	SO_2Ph	MeOH	53% (1.6:1)
6	SO_2Ph	MeCN	64% (7:1)
7	SO_2Ph	PhMe	62% (8:1)
8	SO_2Ph	$PhCF_3$	66% (17:1)
9	SO_2Ph	*p*-xylene	72% (20:1)
10	SO_2Ph	385 nm	17% (20:1)
11	SO_2Ph	450 nm	0%
12	SO_2Ph	0.1 M	80% (20:1)
13	SO_2Ph	0.05 M	84% (20:1)
14	Ts (**3b**)	0.05 M	55%‡ (**4b**)
entry	**PG**	**additive**	**ratio (Pdt:SM) §**
15	COBn (**3n**)	none	1:1.1 (**4n:3n**)
16	COBn	**IX**	3.9:1
17	COBn	**X**	1.15:1
18	COBn	**XI**	0.69:1

Fig. 2. Reaction development. (**A**) Proposed mechanism for piperidine ring contraction. (**B**) Potential undesired side reactivity through Norrish type I and Norrish–Yang cyclization processes. (**C**) Reaction optimization for light-mediated ring contraction. Reactions were performed on a 0.05 mmol scale. Relative stereochemistry is depicted. *Yields were determined by ^{1}H NMR integration using Ph_3CH as an internal standard. †Diastereomeric ratio (d.r.) was determined by ^{1}H NMR integration of resonances corresponding to diastereomers in the crude mixture. ‡Reaction was performed on a 1-g scale in flow, with an isolated yield of major diastereomer reported. §High throughput experimentation (HTE) conducted to identify 3-cyanoumbelliferone. Conversion was determined by liquid chromatography–mass spectrometry analysis.

photoreactive phenyl ketone that could participate in Norrish-type processes (see **VI**). In their studies (Fig. 1D), Suárez and co-workers had achieved success primarily because a photostationary state was reached upon hemiketalization of the α-diketone in the product (*21*, *22*). We envisioned that even though chemical transformation of the photoreactive moiety would not be realized upon product formation in our case, differences in the n→π* photoabsorption profiles of **I** and **V** could arise by virtue of an intramolecular H-bond that is established in **V**. As is well known for spatially forbidden n→π* transitions, increased polarity in the local environment can lead to a hypsochromic (blue) shift for λ_{max} (*24*), which we anticipated would lead to differences in the photoreactivity of **I** and **V**. Additionally, we recognized that other photochemical processes could outcompete our desired ring contraction reaction. For example, although more prominent for dialkyl ketones, Norrish type I C–C bond homolysis from the triplet excited state (i.e., **II**) could lead to the formation of undesired alkyl-acyl radical pairs (see **VII**) (*24*). Another possible complication could arise after 1,5-hydrogen atom abstraction, in which radical recombination of diradical **III** could occur to give the corresponding [3.1.1]-bicycle (**VIII**), i.e., a Norrish–Yang cyclization. We imagined that a Norrish type II C–N fragmentation of **III** (to ultimately afford **V**) would outcompete a Norrish–Yang cyclization on the basis of an anticipated unfavorable conformational bias against ring closure to azetidinol **VIII** and a slower rate for intersystem crossing to the requisite singlet diradical for cyclobutanol C–C bond formation (*25*).

Despite the potential challenges associated with side reactions and overreactions of our phenyl ketone substrate, on the basis of the anticipated differences in absorptivity for **I** and **V,** we reasoned that tuning the wavelength of light could selectively promote the desired reactivity while minimizing the photoreactivity of the resulting products. Gratifyingly, photoirradiation of **3a** using a 400-nm blue LED lamp provided 1,2-disubstituted cyclopentane **4a** in 73% yield at room temperature (Fig. 2C, entry 1). Empirically, the photomediated ring contraction proceeded best with a 400-nm-wavelength light source; longer wavelength irradiation (centered at 450 nm; lower energy) did not result in any conversion, whereas shorter wavelength light sources (e.g., centered at 385 nm; higher energy), which more closely aligned with the calculated and measured absorptivity curves of both product and starting material, led to competing nonspecific oligomerization. Sulfonyl groups on the piperidine nitrogen were especially effective in providing the desired product in good to high yields, presumably enhancing the intramolecular H-bond in **4a** that accounts for its stability under the reaction conditions. Other electron-withdrawing groups on the piperidine nitrogen, such as benzoyl, Boc, or pivaloyl (Fig. 2C, entries 2 to 4), resulted in lower yields of the ring contraction product. The choice of solvent was also critical. Benzene, which can enhance H-bonding in **4a**, gave the highest yields and diastereoselectivity (*26*). Other solvents, such as methanol (Fig. 2C, entry 5), acetonitrile (Fig. 2C, entry 6), toluene (Fig. 2C, entry 7), and trifluorotoluene (Fig. 2C, entry 8), led to diminished yields and lower diastereoselectivity. Given the toxicity concerns of using benzene, we have also identified *p*-xylene as a serviceable solvent alternative (Fig. 2C, entry 9; see the supplementary materials for full solvent

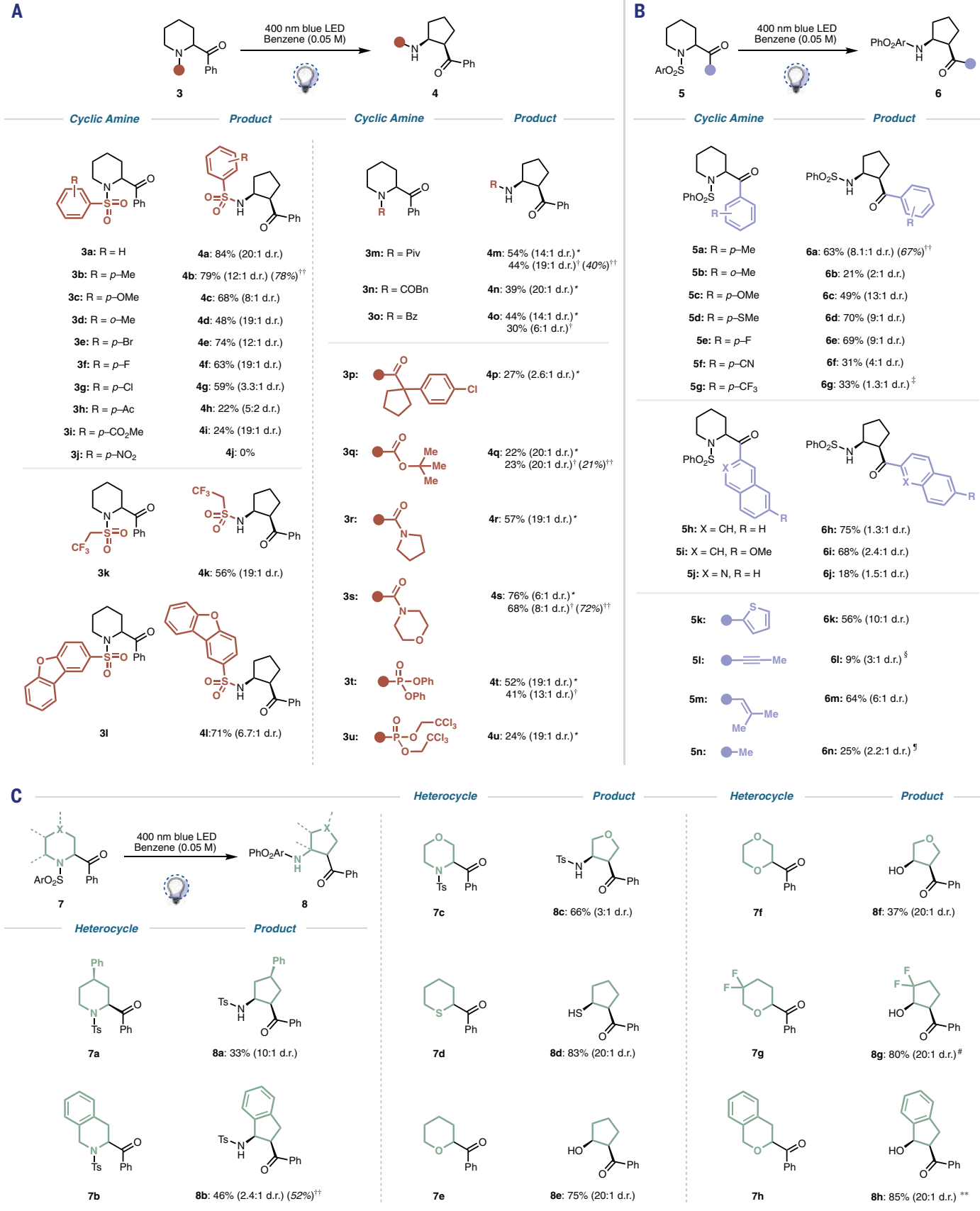


Fig. 3. Scope of substrates in piperidine ring contraction. Reaction conditions: Starting material (0.2 mmol), benzene (0.05 M), 400 nm LED, 18 to 24 hours. Isolated yields reported and relative stereochemistry are shown. Diastereomeric ratio was determined by ^{1}H NMR integration of resonances

corresponding to diastereomers in the crude. (**A**) Scope of protecting groups in ring contraction. *Additive **IX** (30 mol%) was added to the reaction mixture. †0.2 mmol scale trials conducted in the absence of additive **IX**. (**B**) Scope of aryl ketone in ring contraction. ‡Ring-opened product after a subsequent Norrish type II process is also observed. §(With *N*-Ts) Norrish–Yang azetidinol product observed in 12% (19% *brsm*). ¶(With *N*-Ts) Reaction irradiated with a medium-pressure mercury lamp for 24 hours. (**C**) Scope of heterocyclic core in ring contraction. #Reaction conducted on a 0.14 mmol scale. **Reaction conducted on a 0.11 mmol scale. ††Selected examples conducted on a 0.05 mmol scale using *p*-xylene (0.05 M) as a solvent. Yield was determined by ^{1}H NMR integration using Ph_3CH as an internal standard (see the supplementary materials for more details).

screen, and select substrates using *p*-xylene as the solvent). Lowering the concentration of the substrate also led to improvements in yield; a concentration of 0.05 M, which presumably slows the rate of competing unproductive intermolecular side reactions, was determined to be optimal. From these observations, we identified an optimized set of conditions for sulfonyl derivatives of the piperidine substrates, which afforded **4a** in 84% isolated yield and high diastereoselectivity (Fig. 2C, entry 13). The reaction could be performed at gram scale under continuous (flow) conditions (Fig. 2C, entry 14; see the supplementary materials for additional details). For substrates bearing other protecting groups, such as amide and carbamate substrates, we observed reduced conversion to the desired product. This observation was intriguing because, ostensibly, the photoreactive phenyl ketone group in the starting substrates was conserved. We posited that changing the group on nitrogen (e.g., sulfonyl to acyl or carbamoyl) could lead to discrete differences in the triplet state energy of the phenyl ketone. We hypothesized that a photosensitizer or photocatalyst could improve the efficacy of the reaction, and thus we initiated a screen of 46 known photosensitizers and photocatalysts using amide substrate **3n**. 3-Cyanoumbelliferone (**IX**; Fig. 2C, entry 16) emerged as an effective additive that improved the efficiency of the reaction: Piperidine **3m**, **3o**, **3s**, and **3t** were converted to their corresponding cyclopentylamines with improved yields (average of 11% increase, vide infra). The observed effect appeared to be subtle and substrate dependent. Efforts to further understand the role of cyanoumbelliferone **IX** are ongoing.

Scope of the photomediated ring contraction

With optimized conditions in hand, the scope of the visible-light mediated ring contraction was explored (Fig. 3A). The generality of the group on the nitrogen atom was investigated first. Arylsulfonyl-derivatized piperidines featuring electron-withdrawing substituents, such as ketones (e.g., **3h**), esters (e.g., **3i**), and halogens (**3e** to **3g**), led to good to modest yields of the respective ring-contracted products. Likewise, arylsulfonyl-derivatized piperidines bearing electron-donating groups such as ethers (e.g., **3c** and **3d**) were also competent in the ring contraction transformation. The presence of a *p*-nitro substituent (**3j**) resulted in the complete recovery of starting material, likely resulting from photoquenching (*27*–*29*). Although arylsulfonyl substituents on the piperidine nitrogen led to the highest ring contraction product yields, substrates bearing acyl (e.g., **3m** to **3p**), carbamoyl (**3q**), and urea-type groups (e.g., **3r** and **3s**) on the nitrogen also led to successful ring contraction in the presence of cyanoumbelliferone (**IX**) (Fig. 3A).

Phosphoramidite-containing substrates (**3t** and **3u**) were also competent, providing the

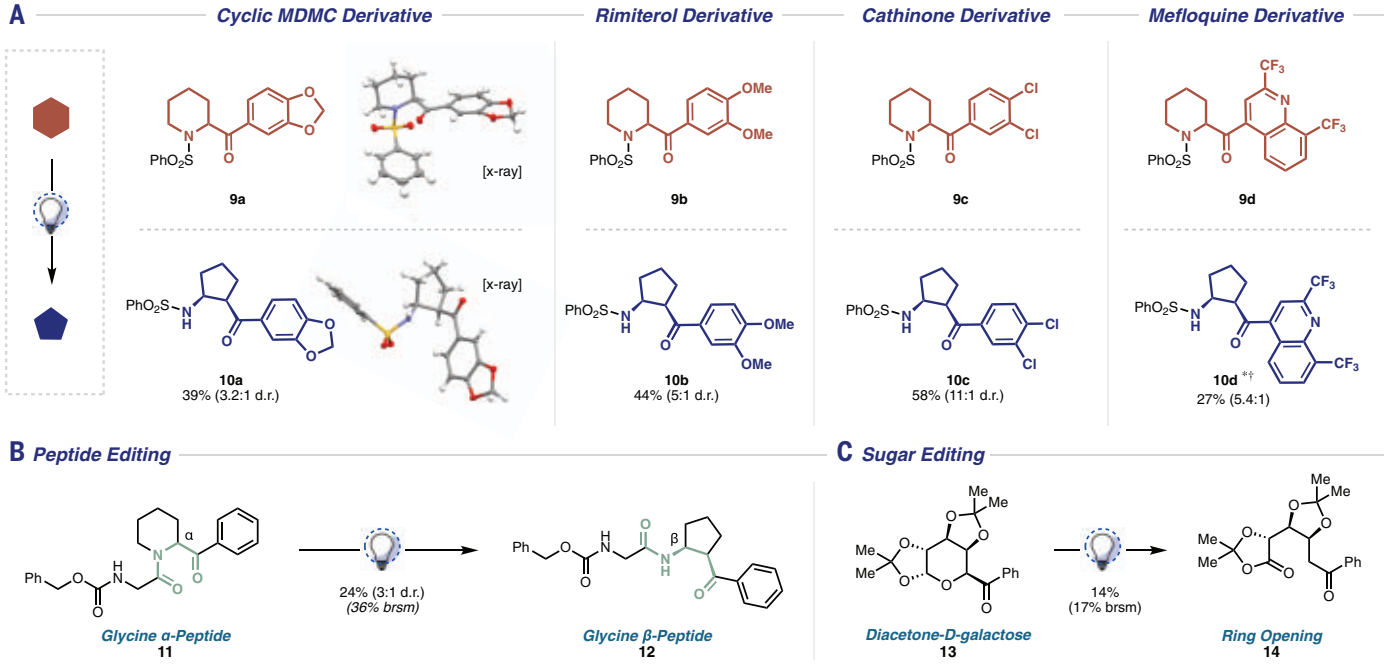


Fig. 4. Applications toward biologically relevant compounds. (**A**) Selected examples of bioactive drug molecule contraction. (**B**) Ring contraction–mediated peptide editing (see the supplementary materials for more details). (**C**) Sugar editing enabled by targeted "digestion." Reaction conditions: Starting material (0.2 mmol), benzene (0.05 M), 400 nm LED, 24 hours. Isolated yields reported and relative stereochemistry are shown. Diastereomeric ratio was determined by ^{1}H NMR integration of resonances corresponding to diastereomers in the crude. *Reaction conducted on a 0.1 mmol scale, where yield was determined by ^{1}H NMR integration using Ph_3CH as an internal standard and d.r. was determined by ^{1}H NMR integration of resonances corresponding to diastereomers in the crude. †Ring-opened product also observed in 4.5% yield (see the supplementary materials for more details).

cyclopentane products in yields comparable to those observed for substrates bearing a sulfonyl group. In these cases, the phosphoramidyl group is easily removed under acidic conditions (*30*). The efficacy of the additive was also found to vary depending on the substrate, with the conversion of **3q** to **4q** being minimally affected upon adding **IX**. Here, the persisting low yield likely involves reduced productive hydrogen atom transfer because of a preference for an unproductive conformation of the starting material (see the supplementary materials for more details).

Next, we investigated the scope of the ring contraction with respect to the nature of the α-aroyl group (Fig. 3B). Here, the reaction was shown to be sensitive to sterics, with *p*-methyl substitution (**5a**) giving rise to higher yields and diastereoselectivity compared with the isomeric *o*-methyl–substituted **5b**. Aryl ketones featuring electron-donating substituents, such as ethers (**5c**) and thioethers (**5d**), led to good yields and diastereoselectivity for the desired product. Halogen-containing α-aryl-ketone **5e** was also competent in the ring contraction chemistry, affording an additional functional handle for modification after ring contraction. By comparison, electron-withdrawing groups such as *p*-CN and *p*-CF_3 (e.g., **5f** and **5g**) led to the formation of the desired product in modest yield and lower diastereoselectivity. This is likely the result of the increased acidity of the α-keto-proton in the product, which could undergo epimerization. On the basis of our calculations, epimerization through a retro–Mannich-Mannich pathway is unlikely (vide infra). Additionally, in the case of substrate **5g**, we observed an ensuing ring opening of the product through Norrish type II cleavage, evidencing the potential overreactivity of the cyclopentane products in some cases (see the supplementary materials for more details). Other aromatic ketones were also examined. For example, extended aromatic systems such as naphthalenyl ketones (**5h**)

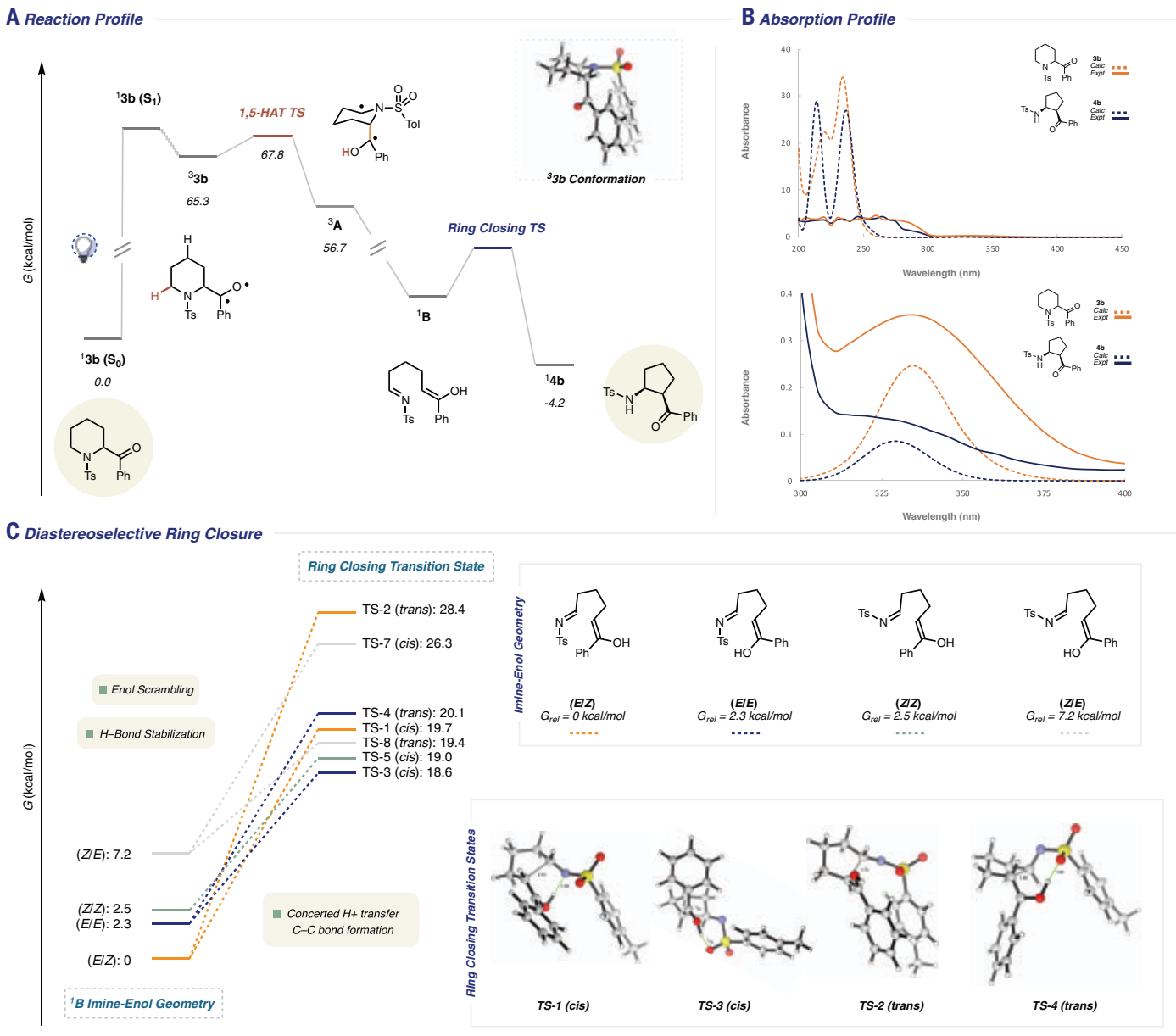


Fig. 5. Computational studies on ring contraction mechanism. (**A**) Reaction profile for the piperidine ring contraction. (**B**) Experimentally and computationally (normalized) determined absorption profiles for the starting material (**3b**) and product (**4b**). (**C**) Imine-enol geometries and transition states calculated for the diastereoselective ring closure.

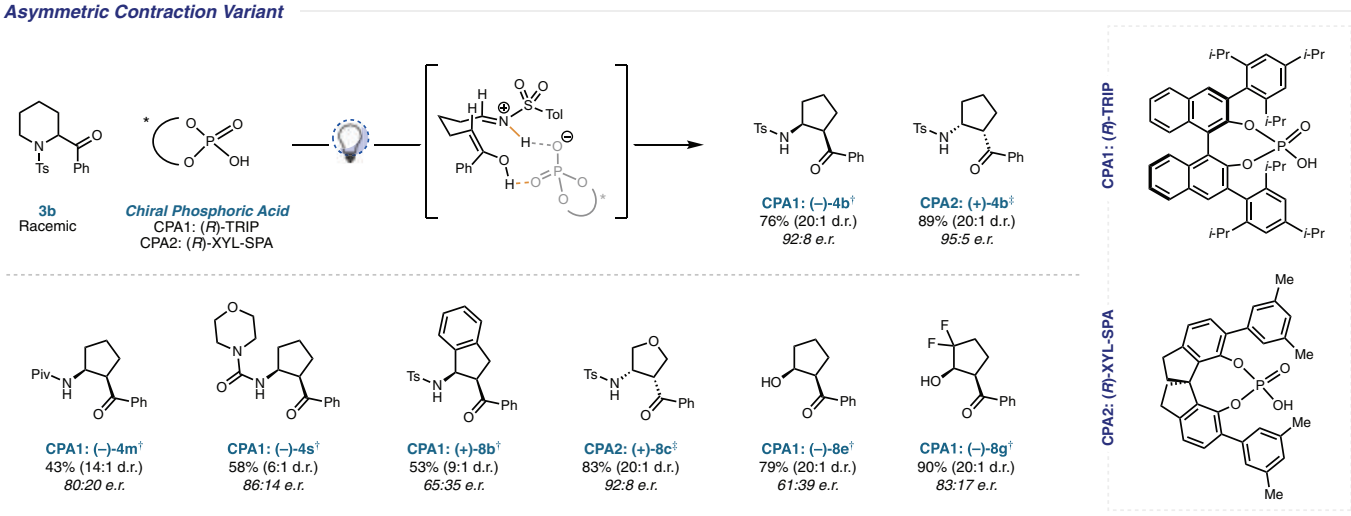


Fig. 6. Development of an asymmetric ring contraction variant. *Reaction conducted on a 0.05 mmol scale in which yields were determined by ^{1}H NMR integration using Ph_3CH as an internal standard and d.r. was determined by ^{1}H NMR integration of resonances corresponding to diastereomers in the crude. For the major diastereomer, e.r. was determined by SFC analysis. †10 mol% (*R*)-TRIP (CPA1) used as chiral phosphoric acid. ‡10 mol% (*R*)-XYL-SPA (CPA2) used as chiral phosphoric acid.

performed comparably, as did heteroaromatics such as thiophenyl ketones (see **5k**). Although the conditions identified were optimized for α-aroyl groups, several substrates bearing non–aryl-substituted ketones also participated successfully in the ring contraction transformation. Specifically, the desired products were observed for an alkynyl C(sp)–bearing substrate (**5l**) and a vinyl C(sp^2)–bearing substrate (**5m**). In the case of **5l**, the substantial decrease in yield can be attributed to a competing Norrish–Yang cyclization process leading to an azetidinol side product (see the supplementary materials for more details). Using more forcing, higher-energy (mercury lamp) irradiation, alkyl C(sp^3)–substituted ketone **5n** was converted to **6n** in 25% yield, highlighting, even in this case, the differential reactivity of the ketone groups in the starting material and product.

Substituted piperidines and several other saturated heterocycles were then examined for their propensity to undergo this type of ring contraction (Fig. 3C). Single substituents at the γ-position were tolerated (see **7a**), imparting stereocontrol. Benzannulated substrates such as tetrahydroisoquinoline **7b** successfully underwent the ring contraction transformation, providing the corresponding amino indane scaffolds (i.e., **8b**) in good yields and under mild conditions. Additionally, upon irradiation of substrates containing a morpholine scaffold, the tetrahydrofuran heterocycle (**8c**) was formed. The ring contraction methodology was also extended beyond azacyclic frameworks to α-acylated thiane and tetrahydropyran derivatives. These were also competent substrates, leading to the formation of cyclopentane thiol and alcohol products (see **8d** to **8h**), respectively. In the case of **7d**, formation of the resulting thiol (**8d**) could be viewed as the unveiling of a covalent modifier upon photoirradiation.

Finally, we turned our attention toward the application of the ring contraction methodology to biologically active small molecules to demonstrate the potential for late-stage derivatization of drug candidates. Upon irradiation, MDMC (**9a**, stimulant), rimiterol (**9b**, bronchodilator), cathinone (**9c**, DAT reuptake inhibitor) (*31*), and mefloquine (**9d**, antimalarial) derivatives underwent contraction to their corresponding cyclopentane isomers (Fig. 4A). In the case of **9d**, a ring-opened aldehyde side product was also observed, potentially arising from hydrolysis of the imine intermediate before Mannich-type ring closure. We speculate that these events were likely due to the electron-deficient bis-trifluoromethylquinoline group in **9d**. The ring contraction transformation was also leveraged in peptide diversification (Fig. 4B). Here, glycine-containing peptide **11** was converted to the corresponding amino cyclopentane (**12**), unveiling an H-bond donor. In this example, irradiation converts the α-peptide grouping to the corresponding β-amino ketone, accomplishing a nonintuitive peptide modification. Prospective applications include unveiling peptide-turn mimics upon irradiation.

Because of the participation of cyclic ether derivatives in these light-mediated ring contractions, we also explored this rearrangement in sugar editing (Fig. 4C). When subjected to 400-nm irradiation conditions, D-galactose–derived bis-acetonide **13** gave isomeric ring-opened product **14**. Here, the enol resulting from the Norrish type II ring opening presumably tautomerizes to the aryl ketone and does not engage the lactone carbonyl group, offering a powerful targeted "digestion" of sugar derivatives.

Computational insight

To gain insight into the proposed mechanism and origin of stereoselectivity for these photomediated ring contractions, we performed a computational study for the reaction of *N*-tosyl piperidine derivative **3b** (Fig. 5A). All of the quantum chemical calculations in the transition state modeling presented were performed using the Gaussian 16 program (*32*). Geometry optimizations and frequency calculations were performed at the M06-2X/6-31 +G** level of theory with the SMD model for implicit solvation by benzene (see the supplementary materials for more details) (*33*–*35*). We initially postulated that the positional selectivity for 1,5-HAT could be attributed to the greater hydricity and lower bond dissociation energy of the α-amino hydrogen atom (Fig. 5A, ^{3}**3b** → ^{3}**A**). Using density functional theory (DFT) calculations, we found that the more hydridic and polarity matched α-piperidinyl hydrogen atom had a HAT transition state 9.0 kcal/mol lower in energy than the potentially competing C–H abstraction at the γ-position (*36*, *37*). The calculations also revealed a conformational preference in the transition state for the N–S bond wherein maximal separation is maintained between the carbonyl and sulfonyl oxygens (see the supplementary materials for more details).

We also reasoned that the diastereoselective formation of the *cis*-disposed amino cyclopentane products (e.g., **4b**) would potentially result from a series of noncovalent interactions such as π-stacking and H-bonding in the transition state for the Mannich-type ring

closure. Computed transition state structures support the Mannich-type cyclization/C–C bond formation proceeding in concert with proton transfer from the enol moiety to the *N*-tosyl group, consistent with the proposal by Suárez and co-workers for the aldol-type cyclization in hexopyranose carbohydrates (*22*). Three scenarios could be envisaged that qualitatively support the experimentally observed diastereoselectivity (Fig. 5C, top right). In the first scenario, only the (*E*/*Z*) imine–enol is productive. This geometry arises when fragmentation of diradical intermediate ^{3}**A** occurs faster than acyl bond rotation (Fig. 5A, see ^{3}**A** orange bond). Alternatively, if interconversion of the (*Z*) and (*E*) forms for both the enol (C=C) and the imine (C=N) bonds is facile, then all four possible imine–enol double-bond geometries can be accessed in the Mannich cyclization. However, the requisite triplet 1,4-diradical conformer that would afford the (*Z*)-imine (Ts-axial) is high in energy, and the contributions from a (*Z*)-imine geometry to the stereochemical outcome are expected to be negligible (see the supplementary materials for a complete discussion). We propose, therefore, that the most likely scenario is one in which the diradical is long-lived enough to allow acyl bond rotation and subsequent sampling of both enol diastereomers (*38*, *39*). In this scenario, only the (*E*/*Z*) and (*E*/*E*) imine–enols are accessible and the energy differences between **TS-1**–**4** would determine the stereochemical outcome. The Boltzmann average of these four transition states predicted a ratio of 14:1 in favor of the *cis* cyclopentane isomer, in good agreement with the experimentally determined 12:1 ratio. The *cis* diastereoselectivity mainly originates from the energy difference between **TS-3** and **TS-4**, which could be rationalized by a shorter and presumably stronger H-bond, as well as a more staggered arrangement of substituents about the forming C–C bond in **TS-3** (Fig. 5C, ^{1}**B** → ^{1}**4b**). The overall transformation of *N*-tosyl piperidine **3b** to either cyclopentane product was calculated to be exergonic (–4.2 kcal mol^{-1} for the **4b** compared with –3.3 kcal mol^{-1} for the *trans*-isomer of **4b**).

Insight into the selective reactivity of the starting material compared with the product under the reaction conditions was also supported by the DFT-calculated and empirically measured absorption profiles of **3b** and **4b**. Even though the calculated λ_{max} values for the starting material and product (Fig. 5B, top) did not differ markedly, we observed secondary absorption peaks associated with the expected n→π* occurring from 300 to 375 nm (Fig. 5B, bottom). Here, a hypsochromic shift was observed for amino cyclopentane product **4b** relative to starting piperidine **3b**, likely accounting for the selective excitation of the starting material. Also observed empirically was an overall decrease in absorptivity for the product (e.g., ε_{340} = 48.0 L cm^{-1} mol^{-1} for **3b** versus ε_{320} = 19.2 L cm^{-1} mol^{-1} for **4b**). Although predictions and rationalizations of photochemical processes tend to focus on λ_{max} values, in our system, irradiation at λ_{max} would have led to indiscriminate reaction of both the starting materials and products. By focusing on the secondary, n→π* absorption region of 300 to 375 nm, we were able to modulate the reactivity of the ketone group that is conserved in both the starting material and product by exploiting subtle differences in the absorbance wavelengths and extinction coefficients (see the supplementary materials for more details) (*40*). The emission spectrum of the 400-nm light source slightly overlaps with the wavelength of light absorbed by the starting material but negligibly with the product. However, the empirically established optimal use of a commercially acquired 400-nm blue LED light source remains to be fully reconciled with the measured absorption values.

Toward a general asymmetric variant

Our insights into the observed diastereoselectivity for these transformations, which arises from highly organized transition states of an achiral imine-enol intermediate (as revealed from our calculations), combined with the successes of other powerful enantioselective photomediated processes (*41*, *42*), inspired us to pursue enantioselective variants. We observed the formation of racemic product **4b** when enantioenriched **3b** was subjected to the ring contraction conditions, confirming that in this formal radical polar crossover the imine–enol intermediate is achiral. Therefore, we could circumvent the inherent challenges associated with stereocontrol of radical intermediates by effecting enantioselective closure of the achiral imine-enol intermediate under a two-electron reaction manifold (*43*). Given the ample literature precedent and predictive models for chiral phosphoric acid–catalyzed reactions of imines, we postulated that rate enhancement of the Mannich step would occur after imine protonation and attendant deprotonation of the enol, leading to an organized transition state favoring attack on one enantioface (see Fig. 6 and the supplementary materials for more details). The combination of H-bonding and ion pairing that would be realized in this case was anticipated to yield enantioenriched Mannich products (Fig. 6) (*44*–*48*).

Consistent with this hypothesis, ring contraction of **3b** with 10 mol% of (*R*)-TRIP (CPA1) as an additive provided (–)-**4b** in 92:8 enantiomeric ratio (e.r.) 84% enantiomeric excess (*ee*) with yields and diastereoselectivity consistent with those obtained under the standard reaction conditions for the formation of the racemate; using the SPINOL-derived phosphoric acid (*R*)-XYL-SPA (CPA2) gave (+)-**4b** in 95:5 e.r. (90% *ee*). Enantioselectivity, albeit modest, was also observed for amide, urea, THIQ, morpholine, and tetrahydropyran derivatives, giving rise to the enantioenriched products **4m**, **4s**, **8b**, **8c**, **8e**, and **8g**, respectively.

Conclusion

Using a Norrish type II reaction, we have established a versatile method for the scaffold remodeling of piperidines as well as other saturated heterocycles. The overall transformation is robust, and the conditions tolerate a wide range of functional groups. Key to the success of these transformations is the "photoprotection" of a pendant ketone group in the product through intramolecular H-bonding, an observation supported by our experimental and computational findings. This reaction has been rendered enantioselective using chiral phosphoric acids.

REFERENCES AND NOTES

1. E. Vitaku, D. T. Smith, J. T. Njardarson, *J. Med. Chem.* **57**, 10257–10274 (2014).
2. A. S. Lawrence, *Amines: Synthesis, Properties and Applications* (Cambridge Univ. Press, 2004).
3. I. Coldham, D. Leonori, *Org. Lett.* **10**, 3923–3925 (2008).
4. K. R. Campos, A. Klapars, J. H. Waldman, P. G. Dormer, C.-Y. Chen, *J. Am. Chem. Soc.* **128**, 3538–3539 (2006).
5. S. Seel *et al.*, *J. Am. Chem. Soc.* **133**, 4774–4777 (2011).
6. W. Chen, L. Ma, A. Paul, D. Seidel, *Nat. Chem.* **10**, 165–169 (2018).
7. S. J. Pastine, D. V. Gribkov, D. Sames, *J. Am. Chem. Soc.* **128**, 14220–14221 (2006).
8. P. Jain, P. Verma, G. Xia, J.-Q. Yu, *Nat. Chem.* **9**, 140–144 (2017).
9. J. B. Roque *et al.*, *ACS Catal.* **10**, 2929–2941 (2020).
10. A. Millet *et al.*, *Chem. Sci.* **4**, 2241–2247 (2013).
11. R. Oeschger *et al.*, *Science* **368**, 736–741 (2020).
12. J. J. Topczewski, P. J. Cabrera, N. I. Saper, M. S. Sanford, *Nature* **531**, 220–224 (2016).
13. K. R. Campos *et al.*, *Science* **363**, eaat0805 (2019).
14. K. A. Tehrani *et al.*, *Tetrahedron Lett.* **41**, 2507–2510 (2000).
15. A. Feraldi-Xypolia, D. Gomez Pardo, J. Cossy, *Chemistry* **21**, 12876–12880 (2015).
16. J. B. Roque, Y. Kuroda, L. T. Göttemann, R. Sarpong, *Nature* **564**, 244–248 (2018).
17. F. Wang, Y. He, M. Tian, X. Zhang, X. Fan, *Org. Lett.* **20**, 864–867 (2018).
18. R. Ling, P. S. Mariano, *J. Org. Chem.* **63**, 6072–6076 (1998).
19. S. Wang, H. Wang, N. Tian, H. Yan, *Tetrahedron Lett.* **65**, 152797 (2021).
20. T. Gees, W. B. Schweizer, D. Seebach, *Helv. Chim. Acta* **76**, 2640–2653 (1993).
21. D. Álvarez-Dorta, E. I. León, A. R. Kennedy, C. Riesco-Fagundo, E. Suárez, *Angew. Chem.* **120**, 9049–9051 (2008).
22. D. Álvarez-Dorta *et al.*, *Chemistry* **19**, 10312–10333 (2013).
23. K. A. Brameld, B. Kuhn, D. C. Reuter, M. Stahl, *J. Chem. Inf. Model.* **48**, 1–24 (2008).
24. E. V. Anslyn, D. A. Dougherty, "Photochemistry" in *Modern Physical Organic Chemistry*, E. V. Anslyn, D. A. Dougherty, Eds. (University Science Books, 2006), pp. 935–1000.
25. M. Oelgemöller, N. Hoffmann, *Org. Biomol. Chem.* **14**, 7392–7442 (2016).
26. C. Reichardt, T. Welton, "Solvent effects on the position of homogeneous chemical equilibria," in *Solvents and Solvent Effects in Organic Chemistry*, C. Reichardt, T. Welton, Eds. (Wiley, 2010), pp. 107–163.
27. Q. Zheng *et al.*, *Chem. Soc. Rev.* **43**, 1044–1056 (2014).
28. R. Dave, D. S. Terry, J. B. Munro, S. C. Blanchard, *Biophys. J.* **96**, 2371–2381 (2009).
29. J. H. M. van der Velde *et al.*, *Faraday Discuss.* **184**, 221–235 (2015).

30. P. G. M. Wuts, T. W. Greene, "Protection for the amino group," in *Greene's Protective Groups in Organic Synthesis*, T. W. Greene, Ed. (Wiley, ed. 4, 2006), pp. 696–926.
31. B. J. Yadav-Samudrala, J. M. Eltit, R. A. Glennon, *ACS Chem. Neurosci.* **10**, 4043–4050 (2019).
32. M. J. Frisch *et al*., "Gaussian 16, Revision A.3" (Gaussian, 2016).
33. Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **120**, 215–241 (2008).
34. A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **113**, 6378–6396 (2009).
35. R. F. Ribeiro, A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **115**, 14556–14562 (2011).
36. A. Padwa, *Tetrahedron Lett.* **5**, 3465–3469 (1964).
37. C. Walling, M. J. Gibian, *J. Am. Chem. Soc.* **87**, 3361–3364 (1965).
38. M. Abe, *Chem. Rev.* **113**, 7011–7088 (2013).
39. S. Muthukrishnan *et al*., *J. Org. Chem.* **75**, 1393–1401 (2010).
40. D. Staveness, J. L. Collins 3rd, R. C. McAtee, C. R. J. Stephenson, *Angew. Chem. Int. Ed.* **58**, 19000–19006 (2019).
41. R. Brimioulle, T. Bach, *Science* **342**, 840–843 (2013).
42. A. Hölzl-Hobmeier *et al*., *Nature* **564**, 240–243 (2018).
43. M. P. Sibi, S. Manyem, J. Zimmerman, *Chem. Rev.* **103**, 3263–3296 (2003).
44. J. M. M. Verkade, L. J. C. Hemert, P. J. L. M. Quaedflieg, F. P. J. T. Rutjes, *Chem. Soc. Rev.* **37**, 29–41 (2008).
45. J. S. Li, Y. J. Liu, S. Li, J. A. Ma, *Chem. Commun.* **54**, 9151–9154 (2018).
46. G. F. Yang *et al*., *J. Org. Chem.* **86**, 5110–5119 (2021).
47. J. P. Reid, L. Simón, J. M. Goodman, *Acc. Chem. Res.* **49**, 1029–1041 (2016).
48. J. P. Reid, M. S. Sigman, *Nature* **571**, 343–348 (2019).

ACKNOWLEDGMENTS

We acknowledge the help and support of the following people from Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA (MSD): S. W. Krska, D. DiRocco, T. W. Lyons, D. Lehnherr, L.-C. Campeau, A. Northrup, N. K. Terrett, E. R. Parmee, and M. H. Kress for support of the Disruptive Chemistry Fellowship program; U. F. Mansoor and E. Corcoran for assistance with flow chemistry; J. Sauri and E. Kwan for assistance in NMR structure elucidation; D. Lehnherr, M. Uehling, D. Kalyani, and S. Lin for assistance with high-throughput experimentation; Y. Jiang, T. Johnson, R. Ayore, and W. Pinto for collecting HRMS data; L. M. Nogle, D. A. Smith, A. Beard, M. Darlak, and M. Pietrafitta for reversed-phase purifications; S. McMinn and L. Nogle for chiral SFC purifications; and N. Sciammetta, A. Musacchio, and E. Corcoran for thorough review of the manuscript and helpful discussions. We also thank A. Reeves for collecting UV-Vis data and J. Roque and J. S. Ham (Berkeley) for their helpful discussions. **Funding:** This work was supported by the Disruptive Chemistry Fellowship program of MSD (to C.S.Y.); the MRL Postdoctoral Research Program of MSD (to M.C.L.); the National Institutes of Health (grant S10OD024998 funding the 600-MHz cryoprobe). R.S. is grateful to the NSF under the CCI Center for Selective C–H functionalization (grant CHE-1700982) and the National Institute of General Medical Sciences (grant R35GM130345A) for partial support of the work reported herein (Berkeley). **Author contributions:** J.J., M.C.L., C.S.Y., and R.S. conceived and designed the experiments reported in this work. J.J., M.C.L., and S.F.K. performed all laboratory experiments. D.A. assisted in structural elucidation. Y.-h.L. designed and completed the DFT calculations. J.J., M.C.L., S.F.K., Y.-h.L., C.S.Y., and R.S. wrote the manuscript. C.S.Y. and R.S. directed the research. **Competing interests:** R.S. is a paid consultant for MSD. The authors declare no other competing interests. **Data and materials availability:** All reported data can be found in the manuscript or the supplementary materials. Requests for additional materials should be directed to the corresponding authors. X-ray crystallographic data for **3a**, **4a**, **4m**, **9a**, **10a**, and (–)-**4b** (CCDC 2061160, 2061161, 2061162, 2061169, 2061170, and 2071240, respectively) are available from the Cambridge Crystallographic Data Centre (https://www.ccdc.cam.ac.uk/).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/1004/suppl/DC1
Materials and Methods
Figs. S1 to S30
Tables S1 to S7
References (*49*–*69*)

25 March 2021; accepted 23 July 2021
Published online 12 August 2021
10.1126/science.abi7183

SUPERCONDUCTIVITY

Field-induced transition within the superconducting state of $CeRh_2As_2$

S. Khim[1]*†, J. F. Landaeta[1]†, J. Banda[1], N. Bannor[1], M. Brando[1], P. M. R. Brydon[2], D. Hafner[1], R. Küchler[1], R. Cardoso-Gil[1], U. Stockert[1], A. P. Mackenzie[1,3], D. F. Agterberg[4], C. Geibel[1], E. Hassinger[1,5]*

Materials with multiple superconducting phases are rare. Here, we report the discovery of two-phase unconventional superconductivity in $CeRh_2As_2$. Using thermodynamic probes, we establish that the superconducting critical field of its high-field phase is as high as 14 tesla, even though the transition temperature is only 0.26 kelvin. Furthermore, a transition between two different superconducting phases is observed in a *c* axis magnetic field. Local inversion-symmetry breaking at the cerium sites enables Rashba spin-orbit coupling alternating between the cerium sublayers. The staggered Rashba coupling introduces a layer degree of freedom to which the field-induced transition and high critical field seen in experiment are likely related.

The vast majority of unconventional superconductors have simple, single-component phase diagrams. This is surprising because the nature of superfluidity in ^{3}He (*1*) and the fact that degeneracies or near-degeneracies can be expected to result from many of the electronic mechanisms for unconventional superconductivity (*2*) suggest that a number of materials should feature temperature–magnetic field phase diagrams with transitions between different superconducting order parameters within the superconducting state. Until now, however, the only stoichiometric superconductor that has been well established to have such a phase diagram at ambient pressure is UPt_3 (*3*–*5*).

Here, we report the discovery of this type of phase diagram in the heavy-fermion material $CeRh_2As_2$. Experimentally, we show that $CeRh_2As_2$ has extremely high superconducting critical fields of up to 14 T despite a superconducting transition temperature T_c of only 0.26 K. Further, when the magnetic field is applied along the crystallographic *c* axis, the superconducting state contains a well-defined internal phase transition at ~4 T, which we identify using several thermodynamic probes. We also suggest that these observations result from physics different from that at play in UPt_3; the key superconducting properties of $CeRh_2As_2$ are likely a manifestation of local inversion-symmetry breaking and consequent Rashba spin-orbit coupling in an overall inversion-symmetric crystal structure (*6*–*13*), a situation for which multiphase superconductivity has been considered in the theoretical literature (*13*–*15*) but not observed in a material thus far. Combined with distinctive normal-state physics that likely also results from the unusual crystalline environment of Ce, our observations suggest that $CeRh_2As_2$ will be a benchmark material in which to study the influence of spin-orbit coupling on electronic mechanisms for unconventional superconductivity.

[1]Max Planck Institute for Chemical Physics of Solids, 01187 Dresden, Germany. [2]Department of Physics and MacDiarmid Institute for Advanced Materials and Nanotechnology, University of Otago, Dunedin 9054, New Zealand. [3]Scottish Universities Physics Alliance, School of Physics and Astronomy, University of St Andrews, St Andrews KY16 9SS, UK. [4]Department of Physics, University of Wisconsin–Milwaukee, Milwaukee, WI 53201, USA. [5]Physik Department, Technische Universität München, 85748 Garching, Germany.
***Corresponding author. Email: elena.hassinger@cpfs.mpg.de (E.H.); seunghyun.khim@cpfs.mpg.de (S.K.)**
†These authors contributed equally to this work.

Heavy-fermion superconductivity in $CeRh_2As_2$

$CeRh_2As_2$ crystallizes in a centrosymmetric tetragonal $CaBe_2Ge_2$-type structure (*16*) (Fig. 1A) in which Ce is alternatively stacked with two different Rh-As blocks along the *c* axis; Rh(1) [As(1)] is tetrahedrally coordinated by As(2) [Rh(2)]. There are two Ce atoms per unit cell. The Ce site lacks local inversion symmetry with the polar C_{4v} point group. The lattice inversion center lies in the middle of the line connecting the two Ce atoms. We believe that this distinctive structural feature plays a central role in the physics of the superconducting state.

The high-temperature magnetization of single-crystalline $CeRh_2As_2$ shows paramagnetic Curie-Weiss behavior with an effective moment of 2.56 bohr magnetons (μ_B) per Ce, corresponding to a Ce^{3+} valence state (Fig. 1B). In the whole temperature range, the *ab* plane magnetization is larger than the *c* axis magnetization by up to a factor of two at low temperature. The resistivity $\rho(T)$ depicted in Fig. 1C displays typical heavy-fermion behavior, with increasing resistivity upon decreasing temperature owing to the Kondo effect. At temperatures below a characteristic local maximum at ~40 K, $\rho(T)$ decreases when the heavy quasiparticle bands are formed by hybridization of local 4*f* electrons with the conduction electrons. The large thermopower $S(T)$ is typical for a Kondo lattice system (Fig. 1D) (*17*). Below 4 K, the specific

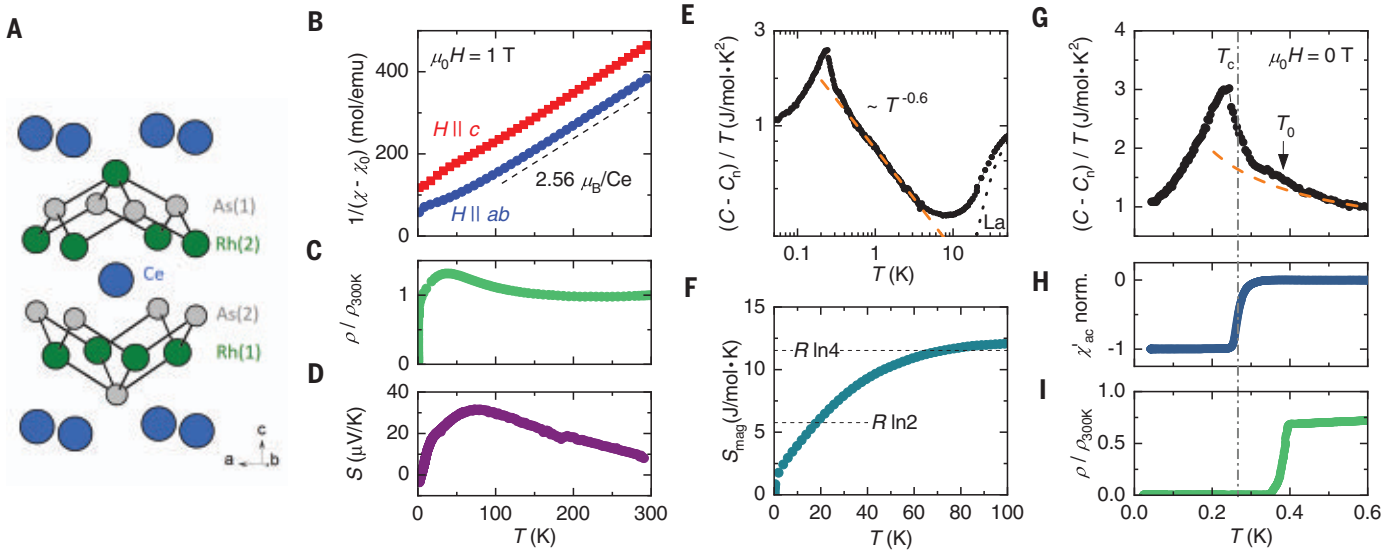


Fig. 1. Crystal structure and heavy-fermion superconductivity in $CeRh_2As_2$. The magnetic field μ_0H = 0 unless indicated otherwise. (**A**) Crystal structure of $CeRh_2As_2$. (**B**) Inverse magnetic susceptibility $\chi(T)$ [after subtracting a temperature (T)–independent contribution χ_0] in μ_0H = 1 T applied in the *ab* plane (blue points) and along the *c* axis (red points). The dashed line denotes the linear slope for the effective moment of Ce^{3+}. emu, electromagnetic unit. (**C**) The resistivity $\rho(T)$ with the current in the *ab* plane, normalized at 300 K. (**D**) The thermopower $S(T)$ with a temperature gradient in the *ab* plane. (**E**) The specific heat $(C - C_n)/T(T)$. A nuclear contribution C_n was subtracted at low T (*17*). The dotted line presents the $LaRh_2As_2$ data used to subtract the phonon contribution. The dashed line represents the power-law T dependence. (**F**) The Ce magnetic entropy $S_{mag}(T)$. (**G** to **I**) Experimental signatures at the superconducting transition T_c and at the transition T_0; see the text for details. (G) Specific heat $(C - C_n)/T(T)$ including the same dashed line as in (E) and transition temperatures as indicated. (H) Normalized ac susceptibility χ'_{ac} (I) The normalized electrical resistivity $\rho(T)$.

heat $C(T)/T$ in Fig. 1E [where the nuclear contribution has been removed (*17*)] increases toward low temperature following a power law with $C/T \propto T^{-0.6}$, suggesting non-Fermi liquid behavior and proximity to a quantum critical point (*18*). C/T reaches a large value of 1 J/mol·K^2 at T = 0.5 K. The Kondo temperature in $CeRh_2As_2$ is between 20 and 40 K, as estimated from the magnetic entropy $S_{mag}(T)$ shown in Fig. 1F (*17*). Of note, S_{mag} monotonically increases to reach the value Rln4 without a plateau at Rln2, where R is the ideal gas constant, suggesting that the two low-lying doublets of the crystal electric field (CEF) are very close in energy. The estimated separation of ~30 K that is comparable with the Kondo energy could lead to a possible quasi-quartet ground state (*17*, *19*). This is a rare example among the tetragonal Ce systems, which usually exhibit a separation of ≳100 K, and again highlights the unusual local Ce environment in $CeRh_2As_2$.

Below 1 K, two anomalies appear in the specific heat, as shown in Fig. 1, E and G. A small hump is visible at $T_0 \approx 0.4$ K, where the data depart from the power-law behavior extrapolated from high temperatures, which is depicted by the dashed line. This departure hints at a phase transition to an ordered state. The large jump below 0.3 K results from the transition to a superconducting state involving the *f* electrons. An equal entropy analysis reveals T_c = 0.26 K and a height of the jump at T_c of $\Delta C/C|_{T_c} \approx 1$, similar to the Bardeen-Cooper-Schrieffer (BCS) value of 1.4. The residual Sommerfeld coefficient $\gamma = C/T$ for $T \to 0$ is possibly a sign of impurities. The diamagnetic drop of the ac susceptibility confirms entry to the superconducting state (Fig. 1H) at a similar T_c for the transition midpoint but a slightly

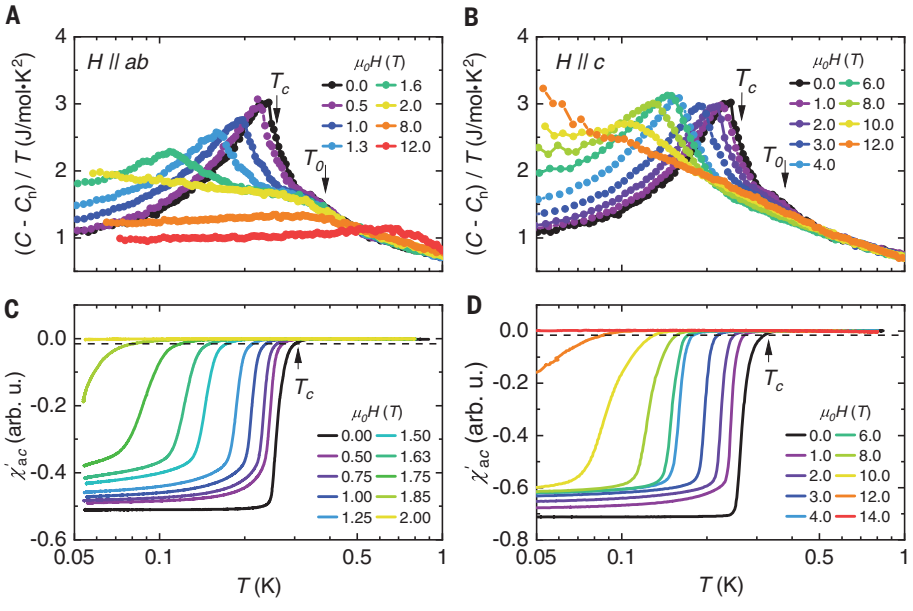


Fig. 2. Evolution of the superconducting transition with magnetic fields. Temperature dependence of the specific heat C/T (**A**) and the real part of the ac susceptibility χ'_{ac} (**C**), respectively, for $H \| ab$. (**B** and **D**) same for $H \| c$. The dashed lines in (C) and (D) indicate the value of χ'_{ac} where the onset temperature T_c is defined. arb. u., arbitrary units.

higher onset temperature. The drop in resistivity takes place at 0.39 K (Fig. 1I). Although this is close to T_0 in zero field, the increase of T_0 with in-plane fields (see the specific heat data at 8 and 12 T in Fig. 2A) shows that T_0 is not associated with superconductivity but likely signals some other kind of order. The origin of the new order is yet to be determined, but the absence of an anomaly in the magnetic susceptibility at T_0 suggests that it might have Ce-4*f* multipolar or nematic character. We ascribe the higher T_c in the resistivity and susceptibility to inhomogeneity in the material, as shown in other heavy-fermion systems (*20*–*22*). These first results indicate that $CeRh_2As_2$ is a heavy-fermion superconductor where the lowest CEF levels are separated by an energy of similar size as the Kondo temperature, both on the order of 30 K. Just before becoming superconducting at low temperature, the system enters a state of unknown origin. For the remainder of this paper, we focus on the extraordinary superconducting properties of $CeRh_2As_2$, as established experimentally using magnetic susceptibility and thermodynamic probes.

Figure 2 shows the temperature dependence of the specific heat C/T (panels A and B) and the magnetic ac susceptibility χ'_{ac} (panels C and D) for different magnetic fields between 0 and 14 T for $H\|ab$ (panels A and C) and $H\|c$ (panels B and D). T_c is defined via the equal entropy method in C/T and at the onset of the susceptibility transition (chosen arbitrarily by the temperature where χ'_{ac} has dropped to the value indicated by the dashed line); T_c shifts down with increasing field. In χ'_{ac}, we observe a relatively strong shift of T_c in a field of 0.1 T that is absent in the specific heat, which is again a sign of nonbulk superconductivity (fig. S7) (*17*). Increasing the field further reduces T_c more slowly. The superconducting transition is completely suppressed down to 0.05 K at magnetic fields of 14 T for $H\|c$ and 2 T for $H\|ab$. We note that, especially for $H\|c$, these are huge critical fields for a superconductor with a T_c of only 0.26 K. For $H\|c$, the temperature sweeps (Fig. 2, B and D) imply a kink in the $T_c(H)$ curve where, above 4 T, the decrease of T_c is slower than below this field.

Two superconducting phases

A pronounced kink in $T_c(H)$ is suggestive of the existence of two superconducting phases. This is confirmed by field sweeps of the ac susceptibility and two separate thermodynamic probes, magnetization and magnetostriction (Fig. 3). Notably, all three provide strong evidence of a phase transition. Below T = 0.2 K, pronounced kinks in all three observables are seen at a characteristic field, H^* ≈ 3.9 T, that is almost temperature independent as it increases from 3.8 T at 0.05 K to 4.0 T at 0.17 K. As shown in Fig. 3A and the inset of Fig. 3B, diamagnetic shielding

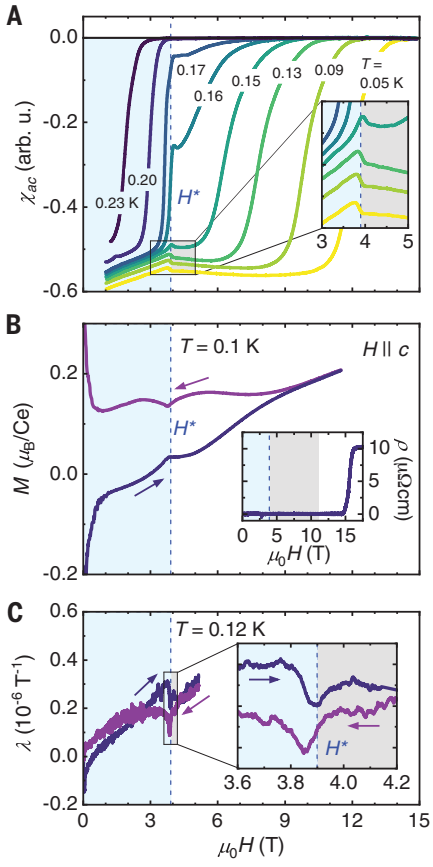


Fig. 3. Phase transition inside the superconducting state for *H*||*c*. (**A**) Absolute value of the magnetic susceptibility χ_{ac} for different temperatures, as indicated. (Inset) Zoom on the transition at H^*. (**B**) Magnetization M at 0.1 K. (Inset) Resistivity at 0.1 K. (**C**) Magnetostriction at 0.12 K. (Inset) Zoom on the transition at H^*. The dashed line is a guide to the eye indicating the H^* transition at ≈3.9 T.

and zero resistivity persist up to the superconducting critical field, further proving that this is a phase transition within the superconducting state. In contrast, field-dependent data for $H\|ab$ show no sign of such a phase transition (fig. S6) (*17*).

Using the values of T_c, H_{c2} [defined in $\chi_{ac}(H)$ at the onset in the same way as in the temperature sweeps], and H^* from our measurements, we show the superconducting phase diagrams of $CeRh_2As_2$ for out-of-plane and in-plane fields in Fig. 4, A and B, respectively. From these phase diagrams, the superconducting critical field can be extrapolated to $H_{c2}(0)$ ≈ 14 T for $H\|c$ and 1.9 T for $H\|ab$. For $H\|c$, two superconducting states appear, labeled as SC1 and SC2, separated by a line that intersects the strong kink in the $H_{c2}(T)$ curve in a multicritical point.

It is useful to estimate the upper critical fields with the Werthamer-Helfand-Hohenberg (WHH) formula $H_{orb} \approx 0.693 T_c \left(-\frac{dH_{c2}}{dT}\right)_{T_c}$, which only uses parameters near T_c where Pauli paramagnetic pair-breaking effects are parametrically suppressed (*23*). Using the large experimental slopes $\left(-\frac{dH_{c2}}{dT}\right)_{T_c} = 97$ T/K for $H\|c$ and 45 T/K for $H\|ab$, this yields H_{orb} ≈ 17 T and H_{orb} ≈ 8 T, respectively. Their anisotropy of a factor of ≈2 reflects the anisotropy of the effective mass, because $H_{orb} \propto m^{*2}$ (*24*). The corresponding BCS coherence lengths $\xi = \sqrt{\Phi_0 / 2\pi H_{c2}(T=0)}$, where Φ_0 is the flux quantum, are accordingly small, lying below 100 Å. The H_{orb} estimates suggest that the upper critical field of SC2 along the *c* axis is not Pauli-paramagnetically suppressed. In contrast, we find that the superconducting state SC1 is strongly Pauli limited, with Pauli critical fields that are enhanced compared with the Clogston-Chandrasekhar limit of ≈0.5 T and 2.5 to 3 times larger for $H\|c$ than for $H\|ab$. This factor represents the scaling factor of the experimental critical field of SC1 for the two magnetic field directions.

The role of spin-orbit coupling

Our key results on the superconducting properties of $CeRh_2As_2$ can be summarized as follows.

1) A large anisotropy in the critical fields, with a *c* axis critical field (14 T) far exceeding the in-plane critical field (1.9 T). Indeed, $\mu_0 H_{c2}/T_c$ for *c* axis fields achieves the highest value thus far observed in Ce-based heavy-fermion superconductors.

2) A superconducting state SC1 from which a second *c* axis field–induced superconducting phase, SC2, appears.

3) The critical fields of SC1 are Pauli limited, in contrast to the critical field for SC2 and $H\|c$, which far exceeds the Pauli field.

4) The *c* axis Pauli field (~5 T) for SC1 is substantially larger than the in-plane Pauli field (1.9 T).

Our first finding, the large anisotropy in the critical fields and large *c* axis critical field, is reminiscent of the noncentrosymmetric heavy-fermion superconductors $CeCoGe_3$, $CeRhSi_3$, and $CeIrSi_3$ (*25*–*29*), whose low crystal symmetry allows Rashba spin-orbit coupling (*29*, *30*). Owing to the broken inversion symmetry in these materials, even-parity (spin-singlet) and odd-parity (spin-triplet) superconducting states are not distinct and are, in general, mixed. These mixed states generically reveal no Pauli paramagnetic suppression for fields along the *c* axis, whereas they do for in-plane fields (*29*, *30*). However, these materials exhibit two important differences with respect to $CeRh_2As_2$: The first is that they do not exhibit multiple superconducting phases, and the second is that inversion symmetry is preserved in $CeRh_2As_2$. These two differences lead to an explanation for our second observation, of superconducting states S1 and S2. Although $CeRh_2As_2$ is centrosymmetric, it is locally noncentrosymmetric,

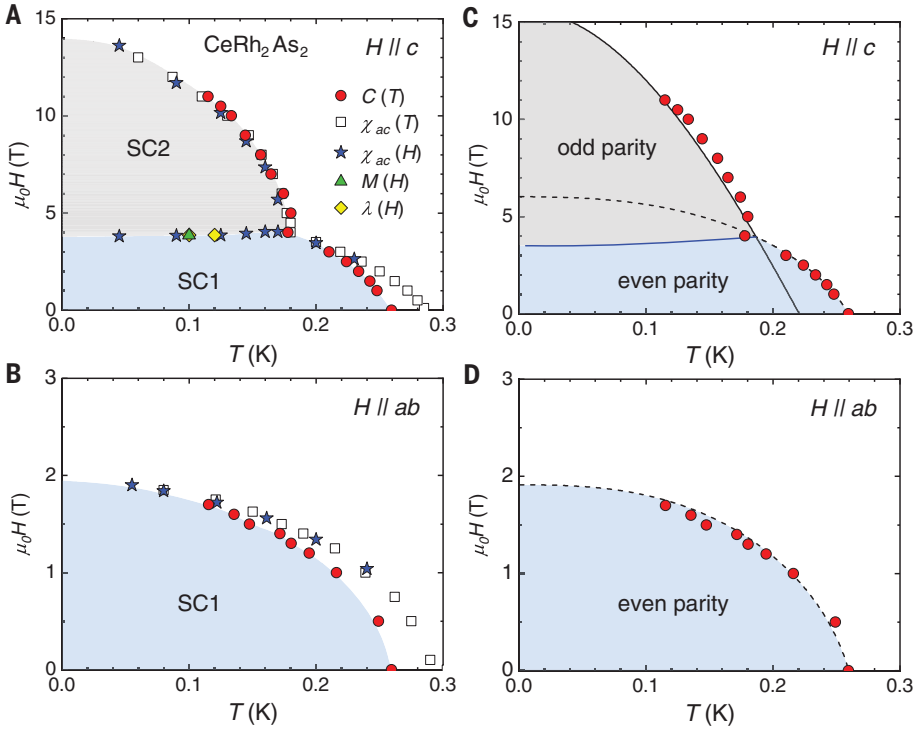


Fig. 4. Superconducting phase diagrams for $CeRh_2As_2$. (**A**) $H \| c$ and (**B**) $H \| ab$. Different symbols are from different experimental probes, as indicated in (A). (**C**) Fits to the upper critical fields for even-parity (dashed line) and odd-parity (solid line) states and a fit to the first-order phase boundary between an even- and odd-parity state (solid blue line). (**D**) Fit to the upper critical field for an even-parity state. For details of the fitting procedure, see (*17*).

with an inversion symmetry linking two noncentrosymmetric Ce-square lattices, each of which has a Rashba interaction. A key feature of the centrosymmetric structure is that even-parity (spin-singlet) and odd-parity (spin-triplet) Cooper pairs are not mixed, such that a phase transition between even- and odd-parity condensates can occur. We argue that this is the case in $CeRh_2As_2$. Indeed, a conceptually similar situation has been considered in models of bilayer materials, where the interplay between an intralayer Rashba interaction and interlayer hopping can lead to a c axis field–driven transition between two superconducting phases (*13*–*15*), similar to that observed here. Below, we show that the structure of $CeRh_2As_2$ allows this bilayer physics to appear in a crystalline setting.

To illustrate the relevance of spin-orbit coupling to determining the key physics of $CeRh_2As_2$ in spite of the presence of global inversion symmetry, we develop a model taking into account the unusual features of its structure. In particular, because Ce 4*f* electrons are key to the heavy quasiparticle bands that give rise to superconductivity, we consider Wannier functions for these bands that are centered on the Ce sites. The Ce atoms sit at sites with a local C_{4v} symmetry, for which electronic states belong to CEF doublets of either Γ_6 or Γ_7 symmetry. A symmetry-based tight-binding Hamiltonian, which takes the same form for either two Γ_6 or Γ_7 doublets, is

$$H_N = t_1[\cos(k_x) + \cos(k_y)] - \mu + \alpha_R \tau_z[\sin(k_x)\sigma_y - \sin(k_y)\sigma_x] + t_{c,1}\tau_x \cos\left(\frac{k_z}{2}\right)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right) + t_{c,2}\tau_y \sin\left(\frac{k_z}{2}\right)\cos\left(\frac{k_x}{2}\right)\cos\left(\frac{k_y}{2}\right) + \lambda\tau_z\sigma_z \sin k_z(\cos k_x - \cos k_y)\sin k_x \sin k_y \quad (1)$$

Here, the σ_i Pauli matrices represent the two Kramer's spin-like degenerate states of the Γ_6 or Γ_7 doublets, and the τ_i Pauli matrices represent the two Ce site degrees of freedom in each unit cell. Given that these two Ce site degrees of freedom are related by inversion symmetry, the τ_z matrix is odd under inversion symmetry (this follows because this matrix changes sign when the two Ce sites are interchanged). Inspection of Eq. 1 shows that a Rashba-like spin-orbit interaction, denoted by the constant α_R, is allowed by symmetry, with the odd inversion symmetry compensated for by the τ_z operator. Equation 1 also reveals an additional Ising-like spin-orbit coupling term denoted by λ, with a $\tau_z\sigma_z$ dependence. This term will be much smaller than α_R because α_R originates from nearest neighbor $(a,0,0)$–type hoppings, whereas λ requires much longer range $(a,2a,c)$–type hoppings; for this reason, we set $\lambda = 0$ for our calculations on $CeRh_2As_2$. The two parameters $t_{c,i}$ correspond to c-axis $(a/2,a/2,c/2)$ hoppings between the two Ce sublattice sites.

We now turn to the superconducting state. Density functional theory (DFT) reveals that the band structure for the conduction electrons is quasi–two dimensional, so it is natural to assume that the quasiparticle interactions that give rise to superconductivity originate in the two-dimensional square Ce layers, which consist of only Ce sites from the same sublattice. So, in terms of τ_i operators, the Cooper pairs can have only a τ_0 or a τ_z dependence. Formally, τ_0 (τ_z) describes Cooper pair wave functions that have the same (opposite) sign on the two Ce sublattice sites. For simplicity, we will assume that each sublattice prefers a spin-singlet s-wave Cooper pair. This is not essential for the arguments presented below, which rely on the τ_i structure of the Cooper pairs; the analysis also applies to a d-wave state such as those commonly found in tetragonal Ce materials. Below, we explicitly consider the even-parity gap function $\Delta_e = \Delta\tau_0$ and the odd-parity gap function $\Delta_o = \Delta\tau_z$. We note that similar gap functions have been discussed in three-dimensional $Cu_xBi_2Se_3$ (*31*), and Δ_o-type gap functions were originally proposed by P. W. Anderson as the generic form of odd-parity superconductivity in heavy-fermion materials (*32*). These results indicate that Δ_e- and Δ_o-type gap functions are stable solutions in general and not only in the quasi–two-dimensional limit considered here.

In (*17*), we carry out a detailed analysis of Eq. 1 on the superconducting state by projecting onto a pseudospin basis. This analysis yields four generic results that apply to all superconducting states described in the previous paragraph: (i) The pairing interactions for Δ_e and Δ_o have the same sign (both are attractive). (ii) Δ_e has a higher transition temperature than Δ_o in zero field. (iii) Δ_e is Pauli suppressed by a c axis field, whereas Δ_o is not (note the suppression of Δ_e by H_z is weaker than the usual paramagnetic suppression, i.e., the Pauli limit is enhanced). A c axis field will therefore induce a first-order phase transition from Δ_e to Δ_o. (iv) For in-plane fields, both Δ_e and Δ_o are Pauli suppressed. These results naturally account for observations 1 and 2 listed at the beginning of this section. We then use a simplified model to fit the data, with the results shown in Fig. 4, C and D. These fits account for observations 1 through 4 and suggest that the situation in $CeRh_2As_2$

is in close correspondence with that described in earlier theoretical work on a model for which, until now, there were no candidate materials (*14*, *15*). Although the core results that we present here are experimental, and we do not claim that our simple model is the only possible explanation for our findings, it is notable that the model succeeds in accounting for all our key observations.

As a final point, we look at the multicritical point in the phase diagram for $H \| c$. In general, thermodynamic considerations forbid that three second-order transition lines meet at a multicritical point (*33*, *34*). However, the phase diagram as experimentally determined here for $H \| c$ is thermodynamically possible if two second-order transition lines and one first-order transition line meet (*34*), and the model suggests that the first-order transition line is the one inside the superconducting state. Several experimental observations point to this scenario as well; an analysis of the slopes of the transition lines and their relation to the size of the specific heat jumps near the multicritical point is consistent with it [fig. S8 and discussion in (*17*)]. Furthermore, the dip in the magnetostriction at H^* corresponds to a steplike change of the sample length and shows hysteresis of ~0.04 T (Fig. 3C). However, these experimental features are extremely small and not confirmed by any other probe, so the experimental evidence for the transition within the superconducting state being first order should not yet be regarded as conclusive.

At present, we cannot exclude the possibility that the phase diagram including the normal state is more complicated. The putative ordered state below $T_0 \approx 0.4$ K also seems to be suppressed near $H \approx 4$ T (Fig. 2B), and the transition line might join the multicritical point as a fourth transition line. Thermodynamically, this would allow the transition within the superconducting state to be second order, and it would place further constraints on the slopes of the lines and the ratios of the specific heat jumps. More generally, it is possible that the change of the superconducting state is influenced by a change in the normal state when the order below T_0 is suppressed. A more detailed study of the specific heat and the magnetocaloric effect would likely be able to resolve the issue.

It is informative to compare our findings with recent developments in UTe_2 (*35*, *36*), wherein multiphase superconductivity has been established in the H–T phase diagram under the application of hydrostatic pressure (*37*) and a splitting of T_c has been reported at ambient pressure (*38*). A substantial body of theoretical work has been done on UTe_2 (*39*–*44*). The prevailing opinion is that the relevant phases are all triplet, and spin fluctuations are thought to be the main driver of the relevant physics. Spin fluctuations are a possible mechanism for superconductivity in $CeRh_2As_2$ as well. Typically, these fluctuations stabilize either even- or odd-parity states, but not both. However, in the unusual electronic environment of $CeRh_2As_2$, Rashba spin-orbit coupling allows both even- and odd-parity states to be stabilized by the same underlying pairing interaction, opening up the possibility of an even-to-odd parity phase transition.

Outlook

Many open questions remain about the precise nature of the superconducting mechanism in $CeRh_2As_2$, including the possibility that the superconductivity in zero applied magnetic field condenses from a normal state that already includes unidentified order. Like its superconductivity, that "hidden" order is probably rooted in the unusual Ce environment. $CeRh_2As_2$ therefore highlights the importance of local symmetry breaking not just for superconductivity but for metallic correlated electronic order as well.

REFERENCES AND NOTES

1. A. J. Leggett, *Rev. Mod. Phys.* **47**, 331–414 (1975).
2. M. Sigrist, K. Ueda, *Rev. Mod. Phys.* **63**, 239–311 (1991).
3. R. A. Fisher *et al.*, *Phys. Rev. Lett.* **62**, 1411–1414 (1989).
4. S. Adenwalla *et al.*, *Phys. Rev. Lett.* **65**, 2298–2301 (1990).
5. R. Joynt, L. Taillefer, *Rev. Mod. Phys.* **74**, 235–294 (2002).
6. M. Shimozawa, S. K. Goh, T. Shibauchi, Y. Matsuda, *Rep. Prog. Phys.* **79**, 074503 (2016).
7. S. L. Wu *et al.*, *Nat. Commun.* **8**, 1919 (2017).
8. K. Gotlieb *et al.*, *Science* **362**, 1271–1275 (2018).
9. H. Ishikawa *et al.*, *Inorg. Chem.* **58**, 12888–12894 (2019).
10. X. Zhang, Q. Liu, J.-W. Luo, A. J. Freeman, A. Zunger, *Nat. Phys.* **10**, 387–393 (2014).
11. L. Yuan *et al.*, *Nat. Commun.* **10**, 906 (2019).
12. M. H. Fischer, F. Loder, M. Sigrist, *Phys. Rev. B* **84**, 184533 (2011).
13. M. Sigrist *et al.*, *J. Phys. Soc. Jpn.* **83**, 061014 (2014).
14. T. Yoshida, M. Sigrist, Y. Yanase, *Phys. Rev. B* **86**, 134514 (2012).
15. D. Maruyama, M. Sigrist, Y. Yanase, *J. Phys. Soc. Jpn.* **81**, 034702 (2012).
16. R. Madar, P. Chaudouet, J. P. Senateur, S. Zemni, D. Tranqui, *J. Less Common Met.* **133**, 303–311 (1987).
17. Materials and methods are available as supplementary materials.
18. A. Steppke *et al.*, *Science* **339**, 933–936 (2013).
19. H. S. Jeevan, C. Geibel, Z. Hossain, *Phys. Rev. B* **73**, 020407 (2006).
20. E. Hassinger *et al.*, *Phys. Rev. B* **77**, 115117 (2008).
21. M. D. Bachmann *et al.*, *Science* **366**, 221–226 (2019).
22. J. F. Landaeta *et al.*, *Phys. Rev. B* **97**, 104513 (2018).
23. N. R. Werthamer, E. Helfand, P. C. Hohenberg, *Phys. Rev.* **147**, 295–302 (1966).
24. J. P. Brison *et al.*, *Physica C Supercond.* **250**, 128–138 (1995).
25. M.-A. Méasson *et al.*, *Physica C Supercond.* **470**, S536–S538 (2010).
26. N. Kimura *et al.*, *Phys. Rev. Lett.* **95**, 247004 (2005).
27. I. Sugitani *et al.*, *J. Phys. Soc. Jpn.* **75**, 043703 (2006).
28. N. Kimura, K. Ito, H. Aoki, S. Uji, T. Terashima, *Phys. Rev. Lett.* **98**, 197001 (2007).
29. E. Bauer, M. Sigrist, Eds., *Non-Centrosymmetric Superconductors: Introduction and Overview*, vol. 847 of *Lecture Notes in Physics* (Springer-Verlag, 2012).
30. M. Smidman, M. B. Salamon, H. Q. Yuan, D. F. Agterberg, *Rep. Prog. Phys.* **80**, 036501 (2017).
31. L. Fu, E. Berg, *Phys. Rev. Lett.* **105**, 097001 (2010).
32. P. W. Anderson, *Phys. Rev. B* **32**, 499 (1985).
33. A. J. Leggett, *Prog. Theor. Phys.* **51**, 1275–1277 (1974).
34. S. K. Yip, T. Li, P. Kumar, *Phys. Rev. B* **43**, 2742–2747 (1991).
35. S. Ran *et al.*, *Science* **365**, 684–687 (2019).
36. D. Aoki, K. Ishida, J. Flouquet, *J. Phys. Soc. Jpn.* **88**, 022001 (2019).
37. D. Aoki *et al.*, *J. Phys. Soc. Jpn.* **89**, 053705 (2020).
38. I. M. Hayes *et al.*, *Science* **373**, 797–801 (2021).
39. K. Machida, *J. Phys. Soc. Jpn.* **89**, 065001 (2020).
40. S. Sundar *et al.*, *Phys. Rev. B* **100**, 140502 (2019).
41. Y. Xu, Y. Sheng, Y.-F. Yang, *Phys. Rev. Lett.* **123**, 217002 (2019).
42. J. Ishizuka, S. Sumita, A. Daido, Y. Yanase, *Phys. Rev. Lett.* **123**, 217001 (2019).
43. J. Ishizuka, Y. Yanase, *Phys. Rev. B* **103**, 094504 (2021).
44. T. Shishidou, H. G. Suh, P. M. R. Brydon, M. Weinert, D. F. Agterberg, *Phys. Rev. B* **103**, 104504 (2021).
45. S. Khim *et al.*, Raw Data for "Field-induced transition within the superconducting state of $CeRh_2As_2$," Max Planck Society (2021); https://dx.doi.org/10.17617/3.72.

ACKNOWLEDGMENTS

We strongly appreciated discussions with D. Cavanagh, O. Erten, M. Fischer, J. S. Kim, D. Möckli, A. Ramirez, B. Schmidt, M. Sigrist, P. Thalmeier, H.-U. Desgranges, H. Rosner, K. Ishida, Y. Yanase, G. Zwicknagl, and S. Wirth. We thank M. König, U. Burkhardt, M. Eckert, and S. Kostmann for EDX measurements. **Funding:** We acknowledge funding from the Physics of Quantum Materials department and the research group "Physics of Unconventional Metals and Superconductors (PUMAS)" of the Max Planck Society. C.G. and E.H. acknowledge support from the German Science Foundation (DFG) through grant GE 602/4-1 Fermi-NESt. P.M.R.B. was supported by the Marsden Fund Council from Government funding, managed by Royal Society Te Apārangi. R.K. is supported by the DFG through project. no. KU 3287/1-1. D.F.A. was supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, under award DE-SC0021971. The research environment in Dresden benefits from the DFG Excellence Cluster Complexity and Topology in Quantum Matter (ct.qmat). **Author contributions:** S.K. and C.G. grew the single crystals. S.K., J.F.L., J.B., N.B., M.B., D.H., R.K., U.S., and E.H. performed the measurements and data analysis. R.C.-G. conducted single-crystal diffraction measurements. P.M.R.B. and D.F.A. constructed the theoretical model. S.K., J.F.L., M.B., D.H., U.S., A.P.M., D.F.A., C.G., and E.H. interpreted the data. All authors were involved in designing the experiment and writing the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data presented in this manuscript are available online at the Open Research Data Repository of the Max Planck Society (*45*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/1012/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 and S2
References (*46*–*62*)

11 September 2020; accepted 23 July 2021
10.1126/science.abe7518

REPORTS

THERMAL TRANSPORT

Low thermal conductivity in a modular inorganic material with bonding anisotropy and mismatch

Quinn D. Gibson[1], Tianqi Zhao[2], Luke M. Daniels[1], Helen C. Walker[3], Ramzy Daou[4], Sylvie Hébert[4], Marco Zanella[1], Matthew S. Dyer[1], John B. Claridge[1], Ben Slater[2], Michael W. Gaultois[1,5], Furio Corà[2], Jonathan Alaria[6]*, Matthew J. Rosseinsky[1]*

The thermal conductivity of crystalline materials cannot be arbitrarily low, as the intrinsic limit depends on the phonon dispersion. We used complementary strategies to suppress the contribution of the longitudinal and transverse phonons to heat transport in layered materials that contain different types of intrinsic chemical interfaces. BiOCl and Bi_2O_2Se encapsulate these design principles for longitudinal and transverse modes, respectively, and the bulk superlattice material $Bi_4O_4SeCl_2$ combines these effects by ordering both interface types within its unit cell to reach an extremely low thermal conductivity of 0.1 watts per kelvin per meter at room temperature along its stacking direction. This value comes within a factor of four of the thermal conductivity of air. We demonstrated that chemical control of the spatial arrangement of distinct interfaces can synergically modify vibrational modes to minimize thermal conductivity.

Lattice thermal conductivity, κ, is a property inherent to all solids, with important technological impact that has driven materials design concepts to maximize the accessible range of values. Electronic devices require high κ to reduce thermal load (*1*) and allotropes of carbon (*2*, *3*) are materials of choice, with isotope control being key to suppress scattering of the lattice vibrations (phonons) that carry heat (*4*). Thermoelectric modules for energy harvesting and thermal barrier coatings for turbine blades require low κ to maintain temperature gradients (*5*). Materials with thermal conductivities lower than that of silica glass (0.9 $W\ K^{-1}\ m^{-1}$), which is used for everyday thermal insulation, are of special interest. The lower limits of κ have been investigated since the publication of theoretical work by Einstein (*6*). The thermal physics is generally understood to become qualitatively different at low κ, with numerous mechanisms appearing in this regime (*7*–*12*). The study of new materials enables improved understanding of previously unexplored atomic arrangements and the arising bonding patterns on κ.

Phonon heat transport can be reduced by decreasing either the phonon scattering length or the phonon group velocity, the long-wavelength limit of which is the material's speed of sound. The scattering length is determined by both intrinsic (phonon-phonon scattering, which is enhanced by anharmonicity) and extrinsic (e.g., defect or boundary scattering) mechanisms, whereas the group velocity is controlled by the phonon dispersion, which is intrinsic to a material because it is defined by the structure and composition. Hence, much research has addressed reducing phonon scattering lengths extrinsically via nanostructuring and defect engineering of materials that are known to have intrinsically low thermal conductivities (*13*, *14*). However, the phonon scattering length has a lower limit of half of the wavelength, below which the vibration can no longer be considered a phonon. This sets an asymptotic high-temperature limit for the thermal conductivity that many materials approach in the high-scattering regime. This limit is determined intrinsically by the entire phonon dispersion regardless of the heat-conduction mechanism (*15*) and sets the scale of accessible κ at all temperatures (*16*). A successful materials design strategy would then reduce this asymptotic limit of κ by engineering the total phonon dispersion defined by the structure at the unit-cell level (*17*, *18*).

Modularity in the buildup of layered crystals presents an opportunity to control materials' properties by careful selection of the interfacial interaction between layers. Analytical models have been developed to describe the impact of the anisotropic bonding strength in layered compounds on their thermal conductivities

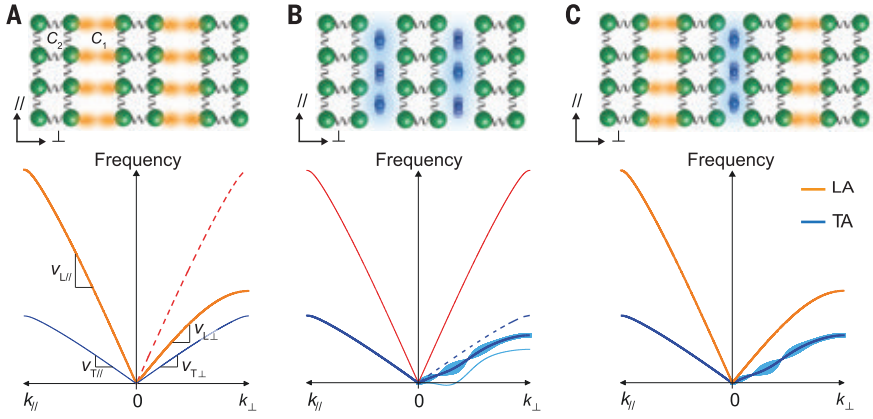


Fig. 1. Chemically compatible interfaces allow synergic phonon dispersion modification and combination driven by bond anisotropy and mismatch. (**A**) Schematic of a material with anisotropic bond contrast produced by stacking of strong and weak bonding interfaces. In the in-plane direction denoted by ||, there are only strong bonds (force constant C_2 denoted by springs). In the out-of-plane direction denoted by ⊥, strong and weak (force constant C_1 denoted by orange shading; $C_1 < C_2$) bonds alternate. The resulting phonon dispersion is shown below, with the LA modes in orange computed from the BvK model ($C_1/C_2 = 0.15$) and a schematic of the TA modes in blue. The slopes are their respective velocities. The red dashed line is the LA mode along $k_\perp$ in the isotropic case where $C_1 = C_2$. The full model for the longitudinal modes is shown in fig. S1. (**B**) Schematic of a material with an incipient in-plane distortion due to an undersized ion layer creating layer-mismatch strain, denoted by the blue shading, and a strong bonding interface, denoted by springs. The resulting schematic phonon dispersion is shown below, with the LA modes in red. The TA modes are in blue, and their line thickness represents the second derivative and thus the anharmonicity, which arises from the proximity to a static distortion. The blue dashed line represents the TA modes in the case with no tendency toward distortion, and the light blue line represents these same modes in the case with a static, dynamically stable in-plane distortion. (**C**) Schematic of a material that combines both anisotropic bond contrast from alternating strong and weak bonding interfaces with an incipient in-plane distortion from layer-mismatch strain. The resulting qualitative phonon dispersion is shown below, demonstrating the integrated effects of all three interfaces in the $k_\perp$ direction.

[1]Department of Chemistry, University of Liverpool, Crown Street, Liverpool L69 7ZD, UK. [2]Department of Chemistry, University College London, 20 Gordon Street, Kings Cross, London WC1H 0AJ, UK. [3]ISIS Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK. [4]Laboratoire CRISMAT, UMR 6508 CNRS, ENSICAEN, UNICAEN, Normandie Université, 6 bd du Maréchal Juin, 14050 Caen, France. [5]Leverhulme Research Centre for Functional Materials Design, The Materials Innovation Factory, University of Liverpool 51 Oxford Street, Liverpool L7 3NY, UK. [6]Department of Physics, University of Liverpool, Oliver Lodge Laboratory, Liverpool L69 ZE, UK.
*Corresponding author. Email: jonathan.alaria@liverpool.ac.uk (J.A.); m.j.rosseinsky@liverpool.ac.uk (M.J.R.)

(*19*, *20*) and to provide a framework for understanding how their asymptotic high-temperature limit falls below the expected isotropic limit through a combination of phonon focusing and first Brillouin zone truncation effects (*16*). We show how different layer sequences can be designed to selectively target longitudinal and transverse phonons and reduce their contribution to the thermal conductivity by generating imbalances in chemical bonding and ion size. The imbalance in chemical bonding substantially modifies the phonon dispersion to reduce heat transport and is structurally compatible with imbalances in ion size that further decrease the thermal conductivity. A designed combination of these interfaces at the unit-cell level in a bulk superlattice material then allows these effects to synergize and affect all transport-active phonon modes along the stacking axis, leading to an extremely low thermal conductivity of 0.10(2) W K^{-1} m^{-1} at room temperature in $Bi_4O_4SeCl_2$, which is among the lowest for any bulk inorganic material and only four times the thermal conductivity of air.

Longitudinal phonons have atomic displacements along the direction of wave propagation and are thus related to compression and expansion modes of the structure, whereas transverse modes have atomic displacements perpendicular to the propagation direction and are related to shear modes. Because these modes originate from different types of atomic displacement, they require distinct design motifs to produce a phonon dispersion that confers low thermal conductivity. In materials with more than one atom per unit cell, phonons have both acoustic (all atoms in the unit cell share a displacement direction) and optical (atoms in the unit cell displace in opposite directions) branches. This leads to transverse acoustic (TA), transverse optical (TO), longitudinal acoustic (LA), and longitudinal optical (LO) phonons. Each acoustic branch has a speed of sound v defined by the group velocity or slope of the dispersion close to the center of the Brillouin zone (Fig. 1A). Because many properties, including thermal conductivity, depend most strongly on the low-frequency regime, the Debye approximation to the dispersion, $\omega(k) = vk$ (where k is the phonon wave vector and ω is the phonon frequency), is often used to derive models describing these quantities. This approximation results in the Debye-Callaway model for the thermal conductivity (*21*), which we refer to as $\kappa_1(T)$ (where T is temperature) (eq. S14), because a single phonon velocity is used; its associated high-temperature asymptotic limit, or minimum thermal conductivity, defined by Cahill and Pohl (*22*), is referred to as κ_{1min}.

We considered the effect of bond contrast at an interface in a layered material to capture the effect of weak interlayer bonds [for example, at a van der Waals (vdW) gap] on the longitudinal phonons that propagate in the out-of-plane (stacking) direction (Fig. 1A). We built a Born von Karman (BvK) model (fig. S1) to describe the longitudinal phonon dispersion on the basis of nearest-neighbor bonding with anisotropic bond contrast, where the weak bonds between the layers with force constant C_1 alternate along the stacking direction with the strong bonds of force constant C_2 ($C_1 < C_2$) that describe the bonds within the layers, in both directions. This model and the resulting phonon dispersion are similar to those described in (*19*)—in our study, we used this phonon dispersion and its parameterization with respect to the bonding interactions to connect the thermal conductivity directly to the bonding via a simple parameterized Debye-Callaway formalism. The in-plane thermal conductivity of the material is described by the κ_1 and κ_{1min} models at any C_1/C_2 ratio, as the in-plane phonon dispersion is entirely determined by C_2.

In the out-of-plane direction, the LA phonon involves primarily interlayer displacements, and its maximum frequency is determined solely by C_1, whereas the LO dispersion depends on both C_1 and C_2, as this phonon involves both intra- and interlayer displacements (fig. S1). As C_1/C_2 decreases, the LA mode becomes more anisotropic with a larger energy gap to the LO mode, which decreases in velocity. The out-of-plane direction then requires a different thermal conductivity model $\kappa_2(T)$ (eq. S24) with two components that describe the LA and LO modes separately to account for their markedly different velocities. The use of two components corrects for the overestimation of heat transfer along the weak interlayer bonds, owing to phonon focusing (*16*, *19*, *20*), and reflects the importance of considering the entire phonon dispersion to describe thermal transport (*15*). To obtain the associated asymptotic high-temperature limit κ_{2min} in the out-of-plane direction, we assume the minimum mean free path is equal to half of the phonon wavelength and use a modified Debye approximation (fig. S22) that is similar in construction to the Debye-Snyder approximation (*18*). κ_{2min} is controlled by C_1/C_2, the out-of-plane speed of sound $v_\perp$, and the atomic number density N and is below κ_{1min} when C_1/C_2 is below 1 (*23*); as C_1/C_2 decreases, so does κ_{2min} (fig. S2). This suggests that anisotropic bond contrast at vdW interfaces in a layered structure can minimize the contribution of longitudinal phonons to thermal transport in the out-of-plane direction.

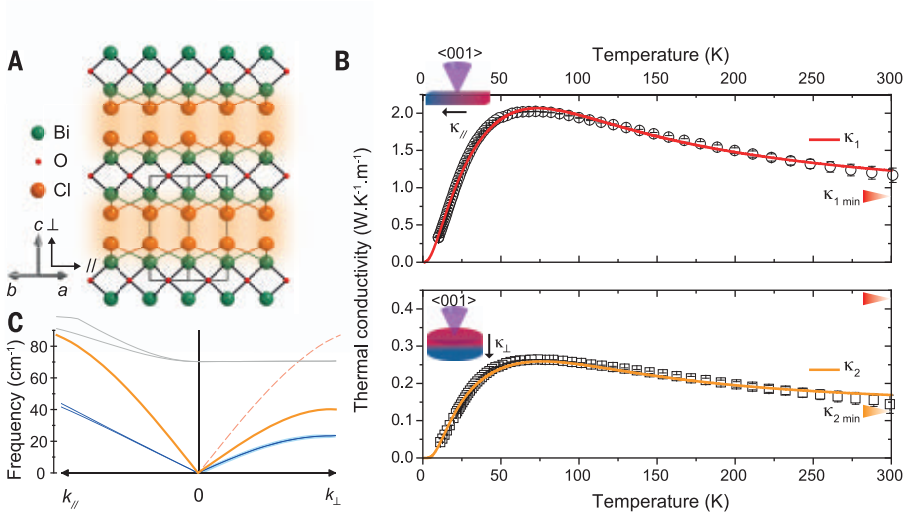


Fig. 2. Longitudinal phonon softening at a vdW interface in BiOCl lowers the out-of-plane thermal conductivity. (**A**) Crystal structure of BiOCl showing the strong bonding and weak vdW bonding (orange shading) interfaces (see fig. S21 for an alternative view). (**B**) Total thermal conductivity of a pressed pellet of BiOCl both parallel (top) and perpendicular (bottom) to the pellet plane. The insets show the orientation of the measurement direction in comparison to the direction of <001> preferred orientation. Fits to the one-component model $\kappa_1(T)$ are shown in red; fits to the two-component model $\kappa_2(T)$ are in orange. The intrinsic asymptotic high-temperature limits κ_{1min} and κ_{2min} are shown by red and orange arrowheads, respectively. Error bars indicate the statistical uncertainty in the measured values. (**C**) Calculated low-frequency phonon dispersion of BiOCl in the in-plane ($k_\parallel$) and out-of-plane ($k_\perp$) directions, highlighting the LA (orange) and TA (blue) modes (the full dispersion is shown in fig. S9). The dashed red line qualitatively illustrates the LA mode along $k_\perp$ in the isotropic case. The LA mode quantitatively matches the BvK model (Fig. 1A) when $C_1/C_2 = 0.054$ (fig. S20). The shading of the TA mode along $k_\perp$ is proportional to the second derivative and, therefore, the anharmonicity.

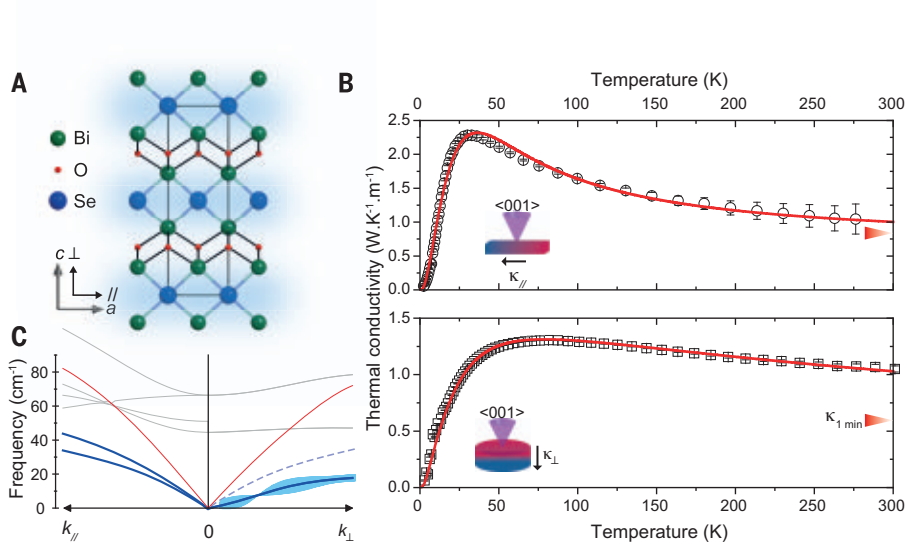


Fig. 3. Transverse phonon anharmonicity at a distortion-prone interface in Bi_2O_2Se. (**A**) Crystal structure of Bi_2O_2Se showing distortion-prone (due to layer-mismatch strain; blue shading) and strong bonding interfaces. (**B**) Total thermal conductivity of a pressed pellet of Bi_2O_2Se both parallel (top) and perpendicular (bottom) to the pellet plane. Fits to the one-component model $\kappa_1(T)$ are shown with solid red lines. Error bars indicate the statistical uncertainty in the measured values. (**C**) Calculated low-frequency phonon dispersion of Bi_2O_2Se in the in-plane ($k_{\parallel}$) and out-of-plane ($k_{\perp}$) directions, highlighting the LA (red) and TA (blue) modes (the full dispersion is shown in fig. S10). The line shading of the TA mode along $k_{\perp}$ is proportional to the second derivative and, thus, the anharmonicity. The dashed blue line qualitatively illustrates the TA dispersion in the case without the tendency toward in-plane distortion.

An alternative strategy can be employed to reduce the contribution of the transverse phonons to the thermal conductivity (Fig. 1B). Materials that are prone to a static distortion producing a structural phase transition often have soft, nonlinear anharmonic TA modes (*24*–*27*). The interface between the undersized Mg layer and the Mg_2Sb_2 layer in Mg_3Sb_2 creates soft anharmonic shear modes (*28*), and the ferroelectric-like incipient lattice distortion that softens phonons in SnSe contributes to its low lattice thermal conductivity (*27*). In a layered material with the stacking axis perpendicular to the basal plane, the out-of-plane transverse phonons exhibit only in-plane displacements. As the chemistry is tuned to vary the energy profile of the in-plane distortion (e.g., by varying the mismatch of size and bonding at the interface between an interstitial ion and a layered building unit), the out-of-plane TA dispersion evolves from linear at small k when there is no distortion to displaying imaginary frequency at the k corresponding to the static displacement of a dynamically stable distortion (*25*), with a softening effect on the transverse phonons in the case of an incipient phase transition (Fig. 1B). The resulting anharmonicity of the out-of-plane TA modes in the latter case will reduce the out-of-plane thermal conductivity by increasing the intrinsic phonon scattering rate. The anharmonicity also leads to a lower velocity for the out-of-plane TA modes ($v_{T\perp}$), which reduces the asymptotic limit of the out-of-plane thermal conductivity.

Combining structurally compatible layered building units that display anisotropic bond contrast and are close to in-plane displacive shearing transitions should integrate their effects on the phonon thermal transport. This requires appropriate chemistry that orders at least three interfaces: a weak and a strong interlayer bonding interface, and an interface that is prone to in-plane distortion. A schematic of such a material is shown in Fig. 1C, together with the targeted phonon dispersion that combines the anisotropic LA mode softening (Fig. 1A) with anisotropic TA mode softening and anharmonicity (Fig. 1B). This highlights the challenge of identifying bonding motifs that allow these characteristic phonon dispersion features of individual modules to be combined without degradation.

To achieve this phonon dispersion in a bulk material, we considered the family of layered $(Bi_2O_2)_mX_n$ materials (where X is any anion but is often a chalcogen or halogen). These materials feature strongly bonded $Bi_2O_2^{2+}$ cationic layers stacked alternately with a wide range of anionic X layers that offer different interfaces, resulting in distinct bonding dimensionalities. Figure 2A shows the crystal structure of BiOCl where the layers are terminated by strong Bi–Cl bonds, such that discrete, strongly bonded $Bi_2O_2Cl_2$ layers are weakly bonded to each other along the c direction by a Cl···Cl vdW interaction (*29*), which can be described as a weak interlayer bond. BiOCl thus has both strong and weak bonding interfaces and should exhibit large anisotropic bond contrast. We measured the thermal conductivities parallel and perpendicular to the plane of a densified pellet of BiOCl (Fig. 2B), which is highly (001) textured (fig. S3) perpendicular to the pellet plane (i.e., in the pressing direction). In this and all materials described here, the measured thermal conductivity is dominated by the lattice contribution, owing to the low electronic conductivities (*23*). The temperature dependence is consistent with crystalline conduction, with a low-temperature peak observed in both directions followed by an inverse temperature dependence toward the asymptotic high-temperature limit. The thermal conductivity we measured parallel to the pellet plane is above κ_{1min}, and its temperature dependence is well described by $\kappa_1(T)$ (Fig. 2B) with appropriate scattering parameters (table S4). Testing other models confirmed that the thermal conductivity should be described in terms of phonon transport (fig. S19). The thermal conductivity we measured perpendicular to the pellet plane is notably different, as the room temperature value of 0.15(2) W K^{-1} m^{-1} is well below κ_{1min} of 0.5 W K^{-1} m^{-1} and continues to decrease upon heating (fig. S7). The thermal conductivity perpendicular to the pellet plane is described well by the two-component model, $\kappa_2(T)$ (Fig. 2B), which was introduced to describe systems with the bonding shown in Fig. 1A, with appropriate scattering parameters (table S3) and κ_{2min} of 0.125 W K^{-1} m^{-1}. The two-component model was needed for the *c*-axis–dominated direction but not the *ab*-plane–dominated direction, which suggests that anisotropic bond contrast in BiOCl induces strong deviation of the longitudinal modes from the Debye model in the out-of-plane direction that suppresses their contribution to the thermal conductivity.

To test the validity of this hypothesis, we calculated the phonon dispersion for BiOCl with first-principles density functional theory (shown up to 100 cm^{-1}; Fig. 2C). The low-frequency dispersion closely resembles that arising from the bonding scheme (Fig. 1A). The predicted anisotropy of the LA mode is evident in the computed speeds of sound (tables S1 and S2), where $v_{L\parallel}/v_{L\perp} = 1.61$ ($v_{L\parallel}$ and $v_{L\perp}$ denote the longitudinal in-plane and out-of-plane speeds, respectively). This LA anisotropy and the large anisotropic energy gap (104 cm^{-1} out of plane) between the LA and first LO phonons in BiOCl show good agreement with the BvK model when $C_1/C_2 = 0.054$ (fig. S20) and arise from the anisotropic bond contrast between the inter- and intralayer displacements that differentiate the LA and LO modes propagating out of the plane (fig. S13). This value of C_1/C_2 is far below 1, confirming that the low value of κ and the applicability

of the κ_{2min} and κ_2 models are due to the structural control of the longitudinal phonons leading to a large anisotropic deviation from the linear Debye model. The transverse modes propagating out of the plane contain only in-plane displacements and thus, to first order, are not affected by the anisotropic bond contrast, which suggests that there is scope to reduce both κ and κ_{2min} further by modifying the phonon dispersion through chemical bonding control beyond the longitudinal modes.

Targeting these out-of-plane transverse phonons requires the introduction of an in-plane structural instability (Fig. 1B) associated with chemical and structural motifs that are compatible with BiOCl. Of the chemically compatible materials with $Bi_2O_2^{2+}$ layers, Bi_2O_2Se has the appropriate bonding (Fig. 3A) to control the transverse modes. A single layer of Se^{2-} anions symmetrically bridges the same $Bi_2O_2^{2+}$ layers to form the anti-$ThCr_2Si_2$ structure, where there is no vdW gap and therefore no associated strong anisotropic bond contrast (*30*). The AM_2X_2 (here A = Se, M = Bi, and X = O) structure field radius ratio (*31*) $f = (r_A)/(r_X + 0.2r_M) = 1.14$ places Bi_2O_2Se in the stability region of the hexagonal anti-$CaAl_2Si_2$ structure adopted by La_2O_2Se, which suggests that Se is undersized in its eight-coordinate environment (with a long 3.4-Å Bi–Se bond) through mismatch to the Bi_2O_2 layer dimensions. The internal lattice strain from this bonding frustration at the bridging Bi_2O_2-Se interface positions the tetragonal structure near a distorted phase (*32*) that is achieved through in-plane static displacements in orthorhombic ultrathin, freestanding Bi_2O_2Se nanosheets (*33*). Therefore, Bi_2O_2Se can be considered to have both a strong bonding interface (within the $Bi_2O_2^{2+}$ unit) and a distortion-prone interface (between the $Bi_2O_2^{2+}$ and Se^{2-} layers).

A pellet of Bi_2O_2Se, which is highly (001) textured (fig. S4), shows low thermal conductivity in both directions [1.0(2) W K^{-1} m^{-1} at room temperature, consistent with previous reports (*34*)] and has the temperature dependence expected for a crystalline material but is, unlike BiOCl, above κ_{1min} (Fig. 3B). It can thus be described by the one-component Debye-Callaway model $\kappa_1(T)$, indicating that this material has weak anisotropic bond contrast ($C_1/C_2 \approx 1$) compared with BiOCl, arising from Se as a bridging ligand in place of terminal Cl. This description is confirmed in the calculated phonon dispersion (Fig. 3C) by an out-of-plane LA mode in Bi_2O_2Se that is much more isotropic ($v_{L\parallel}/v_{L\perp} = 1.10$) and has a higher velocity (3844 m s^{-1}) than in BiOCl (2593 m s^{-1}). The TA modes are controlled by the shear motion of Bi_2O_2 planes against the Se plane, and the small (0.7 cm^{-1}) imaginary frequency in their dispersion (fig. S12) shows the proximity of Bi_2O_2Se to a static in-plane distortion, resulting in low-velocity, anharmonic TA modes propagating out of plane, with the velocity increasing at low frequencies before going to zero at the Brillouin zone edge, leading to an inflection point at 0.025 $Å^{-1}$ (where the second derivative of the dispersion is zero; Fig. 3C). The anharmonicity of the TA phonons is demonstrated by the nonlinearity in the $\omega(k)$ dispersion at low k (*35*–*37*). The second derivative of the dispersion of BiOCl (Fig. 2C) indicates that the dispersion shows little deviation from linearity. Because the LA modes are nearly isotropic, the difference in the temperature dependence of the measured κ between the two directions is attributed to the anharmonicity of the out-of-plane TA modes. The $\kappa_1(T)$ model for both directions could be described using identical values for the extrinsic scattering parameters (table S4). The flattening we observed of the low-temperature peak in the thermal conductivity measured perpendicular to the pellet plane requires an additional intrinsic phonon-phonon scattering term at low frequencies that we ascribed to out-of-plane TA anharmonicity. This anharmonicity also leads to a much lower calculated out-of-plane TA mode velocity in Bi_2O_2Se (836 m s^{-1}) than in BiOCl (1678 m s^{-1}). These observations emphasize that the TA modes in Bi_2O_2Se show the key features of Fig. 1B, leading to a different shape of $\kappa(T)$ for out-of-plane versus in-plane configurations, through mechanisms distinct from, and thus potentially complementary to, the longitudinal mode control achieved in BiOCl.

To create a material in which thermal transport by both longitudinal and transverse modes is suppressed through the combination of anisotropic bond contrast with layer dimension mismatch, we considered the $Bi_2O_2^{2+}$-derived phase $Bi_4O_4SeCl_2$ (Fig. 4A). $Bi_4O_4SeCl_2$ can be described as a bulk superlattice of BiOCl, with contrasting strong and weak bonding interfaces

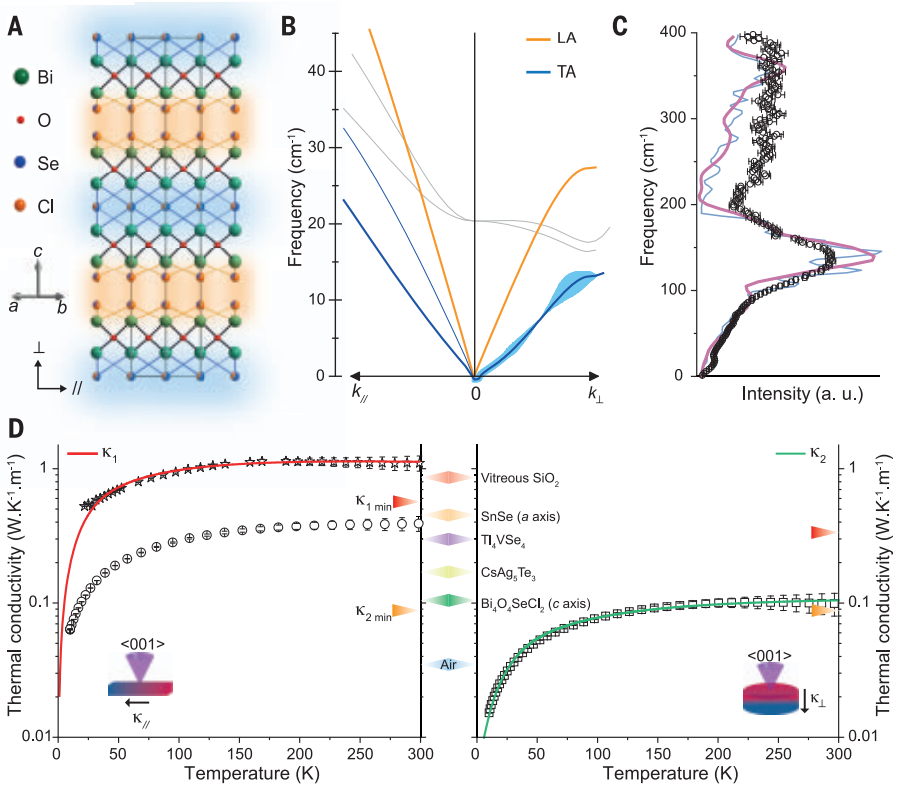


Fig. 4. Extremely low thermal conductivity in $Bi_4O_4SeCl_2$. (**A**) Crystal structure of $Bi_4O_4SeCl_2$, highlighting the three types of interface present: strong bonding, weak vdW bonding, and distortion-prone (see fig. S21 for an alternative view). (**B**) Calculated low-frequency phonon dispersion of $Bi_4O_4SeCl_2$ without Se/Cl site mixing, highlighting the LA and TA modes (full dispersion in fig. S11). (**C**) Comparison of the measured INS (black circles), with the neutron-weighted generalized phonon DOS calculated for $Bi_4O_4SeCl_2$ both with (blue line) and without (magenta line) Se/Cl site mixing. The calculated neutron-weighted phonon DOS does not include the multiple-phonon scattering contribution to the experimental data. Error bars indicate the statistical uncertainty in the measured values. (**D**) Total thermal conductivity of a pressed pellet of $Bi_4O_4SeCl_2$ parallel (left, also including the in-plane single-crystal thermal conductivity) and perpendicular (right) to the pellet plane. In the left panel, the single-crystal in-plane data are shown as stars, and the parallel-to-pellet-plane data are shown as circles. The fits to the one-component model $\kappa_1(T)$ and the two-component model $\kappa_2(T)$ are shown in red and green, respectively. The thermal conductivity is plotted logarithmically for easier comparison both with κ_{1min} and with other low–thermal conductivity bulk inorganic materials (*7*, *8*, *43*, *44*) and air. Error bars indicate the statistical uncertainty in the measured values.

and Bi_2O_2Se containing an undersized bridging Se layer, with its associated distortion-prone interface arising from the in-plane structural instability, and is experimentally stabilized by Se/Cl site mixing (*38*). This extended structure should therefore combine the ingredients of Fig. 1, A and B, to achieve the schematic phonon dispersion presented in Fig. 1C. To evaluate this hypothesis, we calculated the phonon dispersion for $Bi_4O_4SeCl_2$ without site mixing (fig. S11). The low-frequency region of the dispersion (Fig. 4B) displays both the effect of anisotropic bond contrast at the terminally bonded vdW interface on the longitudinal phonons and the creation of soft anharmonic TA phonons through proximity to an in-plane structural instability driven by structural mismatch at the bridging interface (fig. S12), demonstrating the targeted combination of phonon dispersion effects. This integration of both transverse and longitudinal mode control in $Bi_4O_4SeCl_2$ results in a lower average out-of-plane speed of sound than in either BiOCl or Bi_2O_2Se, with an average reduction of 25% compared with BiOCl. In contrast to either end member, the phonon dispersion features both substantially anisotropic LA modes ($^{v_{L\parallel}}/_{v_{L\perp}} = 1.46$) and anharmonic, low-velocity (896 m s^{-1}) out-of-plane TA modes. This reflects the interface synergy in phonon dispersion achieved through the combination of terminal and bridging chemistries in the layer sequence of $Bi_4O_4SeCl_2$.

We measured the neutron-weighted generalized phonon density of states (DOS) of $Bi_4O_4SeCl_2$ with inelastic neutron scattering (INS); it compares well with the calculated neutron-weighted phonon DOS both with and without site mixing (Fig. 4C). The calculated DOS for all materials in our study also agrees well with the thermodynamic parameters extracted from the experimental heat capacity (fig. S18). The Se/Cl site mixing has only a small effect on the calculated DOS, with small changes appearing on optical phonons above 100 cm^{-1}, leaving the low-frequency modes relatively unaffected. Site mixing can lead to increased point-defect scattering, as seen in the increased computed phonon localization in $Bi_4O_4SeCl_2$ with Se/Cl site mixing compared to BiOCl, Bi_2O_2Se, or $Bi_4O_4SeCl_2$ without site mixing (fig. S14); the effect of Se/Cl site mixing is therefore expected to be mostly extrinsic. These observations indicate good qualitative agreement between the calculated and experimental phonon DOS for $Bi_4O_4SeCl_2$, showing that the essential features of the component module phonon dispersions are conserved in the bulk superlattice and should further reduce its thermal conductivity below those of the component motifs.

A pressed, highly (001) textured pellet (figs. S5 and S6) and a single crystal of $Bi_4O_4SeCl_2$ indeed show extremely low and anisotropic thermal conductivity (Fig. 4D) with a characteristic glass-like temperature dependence. We did not observe a low-temperature peak; instead, rather than an inverse decrease, we observed a monotonic increase at high temperature toward the asymptotic limit. This suggests that the heat transport cannot be described using a semiclassical framework but requires consideration of interbranch tunneling and decoherence between vibrational eigenstates (*15*). This transition from crystal to glass-like thermal transport has previously been observed in compounds presenting phonon localization due to site mixing (*39*) and is phenomenologically implemented here by the suppression of the Umklapp scattering contribution in the one- and two-component models. Thus, it does not affect the high-temperature asymptotic limit (as represented by κ_{1min} and κ_{2min}) of these models, as this is governed by intrinsic materials properties controlled by the phonon dispersion. The in-plane thermal conductivity of a $Bi_4O_4SeCl_2$ single crystal is 1.0(1) W K^{-1} m^{-1} at room temperature, remains above κ_{1min} (0.6 W K^{-1} m^{-1}), and can be described with the $\kappa_1(T)$ model with its associated scattering mechanisms (table S4). The thermal conductivity of the textured pellet measured parallel to the pellet plane is very low [0.4(1) W K^{-1} m^{-1}] at room temperature and is slightly lower than the κ_{1min} for that direction because of the contribution of the out-of-plane component from incomplete texturing. The thermal conductivity at room temperature perpendicular to the pellet plane of 0.10(2) W K^{-1} m^{-1} is well below κ_{1min} and may be compared to the 0.026 W K^{-1} m^{-1} thermal conductivity of air. This extremely low value is consistent with the longitudinal phonon dispersion, for which the anisotropic bond contrast effect (Fig. 1A) requires modeling the temperature dependence with the two-component $\kappa_2(T)$ model, as in BiOCl. The complementarity of the transverse phonon mode dispersions through structural mismatch at the bridging interface is highlighted by the further lowering of κ_{2min} compared with BiOCl. The measured asymptotic high-temperature limit of the thermal conductivity for $Bi_4O_4SeCl_2$ is ~30% lower than that for BiOCl (fig. S7), which is very close to the difference in average speed of sound due to the anharmonic TA mode softening in $Bi_4O_4SeCl_2$ (table S1). The thermal conductivity of $Bi_4O_4SeCl_2$ at room temperature is thus among the lowest reported for any bulk inorganic solid (Fig. 4D), achieved by control of the phonon dispersion at the unit-cell level. The distinct bonding motifs in BiOCl and Bi_2O_2Se create low-velocity longitudinal and transverse phonons, respectively. Their complementary chemistries then allow their integration as modules to afford the single-phase superlattice $Bi_4O_4SeCl_2$, where their individual properties are compatibly combined to enhance performance.

These two modules represent the broad classes of vdW materials and materials close to electronically driven structural instabilities, indicating diverse functional opportunities through combination of two or more such units. For example, the single property of thermal conductivity is addressed in $Bi_4O_4SeCl_2$, introduction of electronic carriers complementary to these phonons in a module selected to optimize the power factor (*40*–*42*) would enhance thermoelectric performance, and appropriate electron-phonon coupling across modules might open access to superconductivity. This potential for multiple-property optimization illustrates how synergy between modular units with compatible bonding can enable chemical generation and control of function.

REFERENCES AND NOTES

1. A. L. Moore, L. Shi, *Mater. Today* **17**, 163–174 (2014).
2. S. Berber, Y.-K. Kwon, D. Tománek, *Phys. Rev. Lett.* **84**, 4613–4616 (2000).
3. G. A. Slack, *J. Phys. Chem. Solids* **34**, 321–335 (1973).
4. K. Chen *et al.*, *Science* **367**, 555–559 (2020).
5. G. J. Snyder, M. T. Agne, R. Gurunathan, *Natl. Sci. Rev.* **6**, 380–381 (2019).
6. A. Einstein, *Ann. Phys.* **35**, 679–694 (1911).
7. H. Lin *et al.*, *Angew. Chem. Int. Ed.* **55**, 11431–11436 (2016).
8. S. Mukhopadhyay *et al.*, *Science* **360**, 1455–1458 (2018).
9. R. Pohl, *Phys. Rev. Lett.* **8**, 481–483 (1962).
10. B. Taylor, H. Maris, C. Elbaum, *Phys. Rev. Lett.* **23**, 416–419 (1969).
11. S. Lee *et al.*, *Nat. Commun.* **5**, 3525 (2014).
12. M. T. Agne, R. Hanus, G. J. Snyder, *Energy Environ. Sci.* **11**, 609–616 (2018).
13. X. Qian, J. Zhou, G. Chen, *Nat. Mater.* 10.1038/s41563-021-00918-3 (2021).
14. C. Chiritescu *et al.*, *Science* **315**, 351–353 (2007).
15. M. Simoncelli, N. Marzari, F. Mauri, *Nat. Phys.* **15**, 809–813 (2019).
16. Z. Chen, C. Dames, *Appl. Phys. Lett.* **107**, 193104 (2015).
17. Z. Chen, X. Zhang, S. Lin, L. Chen, Y. Pei, *Natl. Sci. Rev.* **5**, 888–894 (2018).
18. E. S. Toberer, A. Zevalkink, G. J. Snyder, *J. Mater. Chem.* **21**, 15843–15852 (2011).
19. Z. Wei, Y. Chen, C. Dames, *Appl. Phys. Lett.* **102**, 011901 (2013).
20. A. J. Minnich, *Phys. Rev. B* **91**, 085206 (2015).
21. J. Callaway, *Phys. Rev.* **113**, 1046–1051 (1959).
22. D. G. Cahill, R. Pohl, *Solid State Commun.* **70**, 927–930 (1989).
23. Materials and methods are available as supplementary materials.
24. D. J. Singh, *Phys. Rev. B* **53**, 176–180 (1996).
25. D. Sarkar *et al.*, *J. Am. Chem. Soc.* **142**, 12237–12244 (2020).
26. A. K. Tagantsev *et al.*, *Nat. Commun.* **4**, 2229 (2013).
27. C. W. Li *et al.*, *Nat. Phys.* **11**, 1063–1069 (2015).
28. W. Peng, G. Petretto, G.-M. Rignanese, G. Hautier, A. Zevalkink, *Joule* **2**, 1879–1893 (2018).
29. F. A. Bannister, *Mineral. Mag. J. Mineral. Soc.* **24**, 49–58 (1934).
30. H. Boller, *Monatsh. Chem.* **104**, 916–919 (1973).
31. W. Peng, S. Chanakian, A. Zevalkink, *Inorg. Chem. Front.* **5**, 1744–1759 (2018).
32. M. Wu, X. C. Zeng, *Nano Lett.* **17**, 6309–6314 (2017).
33. T. Ghosh *et al.*, *Nano Lett.* **19**, 5703–5709 (2019).
34. P. Ruleova *et al.*, *Mater. Chem. Phys.* **119**, 299–302 (2010).
35. A. Askar, *Proc. R. Soc. London Ser. A* **334**, 83–94 (1973).
36. R. Zivieri, F. Garesci, B. Azzerboni, M. Chiappini, G. Finocchio, *J. Sound Vibrat.* **462**, 114929 (2019).
37. N. W. Ashcroft, N. D. Mermin, in *Solid State Physics* (Holt, Rinehart and Winston, 1976).
38. Q. D. Gibson *et al.*, *J. Am. Chem. Soc.* **142**, 847–856 (2020).
39. L. M. Daniels *et al.*, *Energy Environ.* **10**, 1917–1922 (2017).
40. S. Roychowdhury *et al.*, *Science* **371**, 722–727 (2021).
41. J. R. Sootsman *et al.*, *Angew. Chem.* **47**, 8618–8622 (2008).
42. L. Wu *et al.*, *NPG Asia Mater.* **9**, e343 (2017).
43. J. J. Freeman, A. C. Anderson, *Phys. Rev. B* **34**, 5684–5690 (1986).
44. L.-D. Zhao *et al.*, *Science* **351**, 141–144 (2016).

45. Q. D. Gibson *et al*., Low thermal conductivity in a modular inorganic material with bonding anisotropy and mismatch, DataCat: The Research Data Catalogue (2021); https://datacat.liverpool.ac.uk/id/eprint/1239.

ACKNOWLEDGMENTS

Funding: This project has received funding from the Engineering and Physical Sciences Research Council (EPSRC; grant EP/N004884). We thank the UK STFC for access to the Merlin instrument at ISIS Spallation Neutron Source (Xpress proposal 1990010). The x-ray diffraction facility used to collect pole figures was supported by EPSRC (grant EP/P001513/1). We acknowledge the use of ARCHER UK National Supercomputing Service for this work (EPSRC grants EP/L000202, EP/R029431, and EP/T022213). J.A. thanks the Royal Society for support under the International Cost Exchange IEC\R2\170036. S.H. thanks the CNRS for support through Thermospin PRC. M.J.R. thanks the Royal Society for a Research Professorship. **Author contributions:** Q.D.G., J.A., and M.J.R. developed the concept and wrote the first draft of the manuscript. Q.D.G., M.W.G., L.M.D., J.A., and M.J.R. wrote the final draft of the manuscript. Q.D.G. synthesized the materials. L.M.D. carried out materials processing. Q.D.G., L.M.D., S.H., R.D., and J.A. carried out physical property measurements. H.C.W. performed the INS measurements. M.Z. performed the scanning electron microscopy experiments. Q.D.G. and J.A. analyzed and modeled the physical properties. T.Z., B.S., M.S.D., and F.C. carried out the computational work. J.B.C. contributed to the interpretation of the real and reciprocal space models. All authors were involved in discussions and evaluation of drafts during the writing process. M.J.R. directed the research. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All underlying data are available at the University of Liverpool Research Data Catalogue (*45*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/1017/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 to S4
References (*46*–*59*)

9 April 2021; accepted 25 June 2021
Published online 15 July 2021
10.1126/science.abh1619

APPLIED PHYSICS

Highly conductive and elastic nanomembrane for skin electronics

Dongjun Jung[1,2]†, Chaehong Lim[1,2]†, Hyung Joon Shim[1,2]†, Yeongjun Kim[1,2], Chansul Park[1,2], Jaebong Jung[3], Sang Ihn Han[1,2], Sung-Hyuk Sunwoo[1,2], Kyoung Won Cho[1,2], Gi Doo Cha[1,2], Dong Chan Kim[1,2], Ja Hoon Koo[1], Ji Hoon Kim[3], Taeghwan Hyeon[1,2]*, Dae-Hyeong Kim[1,2,4]*

Skin electronics require stretchable conductors that satisfy metallike conductivity, high stretchability, ultrathin thickness, and facile patternability, but achieving these characteristics simultaneously is challenging. We present a float assembly method to fabricate a nanomembrane that meets all these requirements. The method enables a compact assembly of nanomaterials at the water–oil interface and their partial embedment in an ultrathin elastomer membrane, which can distribute the applied strain in the elastomer membrane and thus lead to a high elasticity even with the high loading of the nanomaterials. Furthermore, the structure allows cold welding and bilayer stacking, resulting in high conductivity. These properties are preserved even after high-resolution patterning by using photolithography. A multifunctional epidermal sensor array can be fabricated with the patterned nanomembranes.

Skin electronics are a set of skin-mounted devices whose mechanical properties are comparable to those of human skin (*1*, *2*). They have been deployed in various applications such as biomedical devices (*3*, *4*), human–computer interfaces (*5*, *6*), and virtual or augmented reality devices (*7*). One of the vital components for skin electronics is the ultrathin elastic conductor (*8*, *9*), and conductive elastic nanocomposites have been considered as a feasible candidate (*10*, *11*). Despite previous research efforts (*12*, *13*), however, critical challenges still remain, including metallike conductivity (*14*, *15*), high stretchability (*16*, *17*), ultrathin thickness (*18*, *19*), and facile patternability (*20*, *21*). Typically, there are trade-offs between these properties (*22*), and these goals have not yet been achieved at the same time (*23*).

[1]Center for Nanoparticle Research, Institute for Basic Science (IBS), Seoul 08826, Republic of Korea. [2]School of Chemical and Biological Engineering, and Institute of Chemical Processes, Seoul National University, Seoul 08826, Republic of Korea. [3]School of Mechanical Engineering, Pusan National University, Busan 46241, Republic of Korea. [4]Department of Materials Science and Engineering, Seoul National University, Seoul 08826, Republic of Korea
***Corresponding author. Email: thyeon@snu.ac.kr (T.H.); dkim98@snu.ac.kr (D.-H.K.)**
†These authors contributed equally to this work.

We present a float assembly method to fabricate a conductive and elastic nanomembrane. This method enables close packing of nanomaterials as a monolayer at the water–oil interface and thus fabrication of a nanomembrane with a cross-sectional structure in which closely assembled metal nanomaterials are partially embedded in an ultrathin elastomeric membrane. Figure 1 presents schematic illustrations that describe the fabrication of a nanomembrane. It consists of three steps, including dropping the nanocomposite solution on deionized water, adding a surfactant, and drying the solvent. The nanocomposite solution is composed of nanomaterials, elastomer dissolved in the water-immiscible solvent, and ethanol. This process can be applied to various nanomaterials, including silver nanowires (Ag NWs), silver-gold core-shell nanowires (Ag-Au NWs), silver nanoparticles (Ag NPs), and gold nanoparticles (Au NPs), and various elastomers such as poly(styrene-ethylene-butylene-styrene) (SEBS), thermoplastic polyurethane (TPU), and poly(styrene-isoprene-styrene) (SIS).

The fabrication process begins with injection of the nanocomposite solution onto water (Fig. 1A). A representative case using Ag NWs and SEBS in toluene is demonstrated. When a solution droplet touches the water surface, separation of the solution occurs: Water-miscible ethanol dissolves into water, whereas water-insoluble elastomer (i.e., SEBS) and water-immiscible solvent (i.e., toluene) stay on the water (fig. S1, A to E). This forms an interface between water (including ethanol) and toluene (including SEBS). NWs with amphiphilic ligands [e.g., polyvinyl pyrrolidone (PVP)] settle at the interface, which stabilizes the system by lowering the interfacial energy between water–ethanol and toluene–SEBS. If the ligand is more hydrophilic or hydrophobic than PVP, NWs become located in either water or toluene, respectively (fig. S1F).

Dissolution of ethanol in water decreases local surface tension, which results in a circular surface tension gradient near the droplet (fig. S2, A to C). This gradient induces Marangoni flow from the center to the boundary, which drags the floating mass including toluene, SEBS, and NWs. The boundary expands as ethanol spreads out. Thus, the mass moves until it reaches the previously transferred mass, and they merge (fig. S2, D to F). An overview of this assembly process is shown in Fig. 1B. Because Marangoni flow transfers the mass rapidly, NWs are assembled as a monolayer, and their agglomeration is suppressed (fig. S2, G and H). Without ethanol (i.e., without Marangoni flow), however, NWs aggregate (fig. S2, I and J), despite using the same process.

As the solution injection continues, the assembled mass is accumulated from the edge of the container to the center, which eventually covers the entire water surface (movie S1 and fig. S3A). At this stage, NWs on water are not fully packed (Fig. 1C). Thus, a few drops of a surfactant (e.g., IGEPAL CO-520) are added at the center (Fig. 1D), which pushes the mass outward (Fig. 1E, movie S2, and fig. S3B). As a result, NWs can be more closely packed with the addition of surfactant (fig. S4, A to C).

The solvent evaporates within 5 min at room temperature, and an elastic conductive nanomembrane, which is a monolayer of assembled NWs partially embedded in an ultrathin elastomer membrane, remains on the water (Fig. 1F). This method is scalable, and a

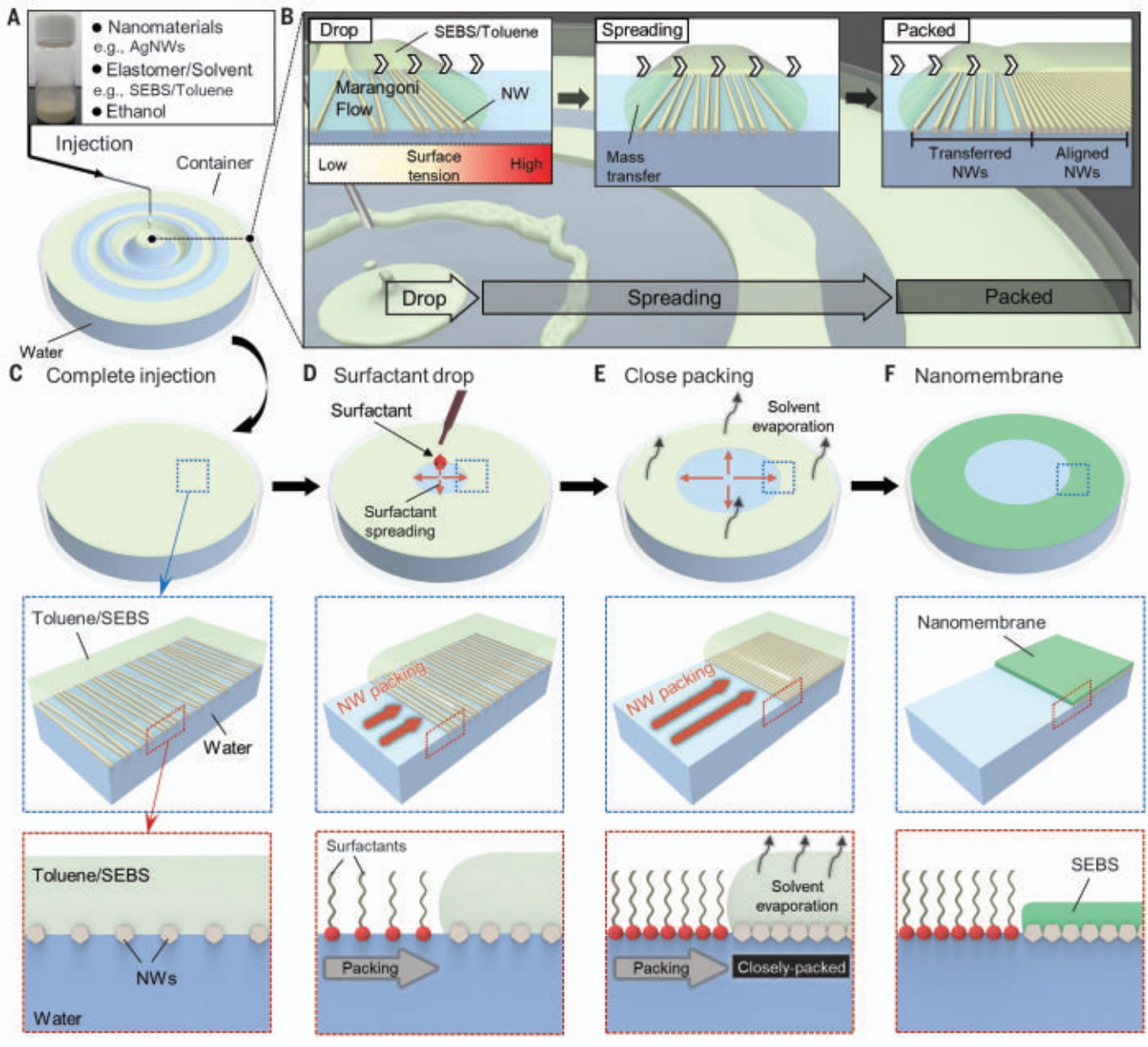


Fig. 1. Fabrication of a highly conductive and stretchable nanomembrane using a float assembly method. (**A** and **B**) Schematic illustrations of the solution injection process. (A) The fabrication begins with injection of the nanocomposite solution onto water. The solution consists of nanomaterials, water-insoluble elastomer dissolved in a water-immiscible solvent, and ethanol. (B) The mass of the nanocomposite solution spreads out along the water surface as a consequence of Marangoni flow, resulting in the monolayer assembly of NWs. (**C** to **F**) Schematic illustrations that show the close-packing process of NWs in a macroscopic view (top), a magnified view inside the indexed blue box (middle), and a more magnified cross-sectional view inside the indexed red box (bottom). (C) The assembled mass covers the entire water surface after the solution injection process. (D) A few drops of a surfactant are added at the center. (E) The surfactant pushes the mass (i.e., NWs, elastomer, and solvent) outward. The solvent evaporates quickly at room temperature. (F) A monolayer of assembled NWs partially embedded in an ultrathin elastomer matrix is left on water.

large nanomembrane can be fabricated. Despite the large size of the container (a diameter of 30 cm), the thickness and structure are uniform (fig. S4, D to F). The nanomembrane can be transferred onto various substrates, such as a wafer, a plastic substrate, or even an elastomeric substrate, for further processing.

In general, loading a high amount of metallic fillers in elastomeric nanocomposites leads to high conductivity (*24*). However, this renders the nanocomposite stiff and brittle, and it eventually loses the original elasticity (*25*). The viscous property of the elastomeric nanocomposite solution hinders its facile fabrication into an ultrathin layer because of the friction force on most solid-state substrates. The closely packed monolayer assembly of NWs on water, however, enables fabrication of an ultrathin nanomembrane that features both a large weight fraction of NWs (>80 wt %) and a high elasticity.

Such material properties are attributed to its cross-sectional structure, which derives from the monolayer assembly of NWs at the interface between water and oily solvent. Half of the NWs are fixed by the elastomer membrane, whose cross-sectional structure resembles teeth

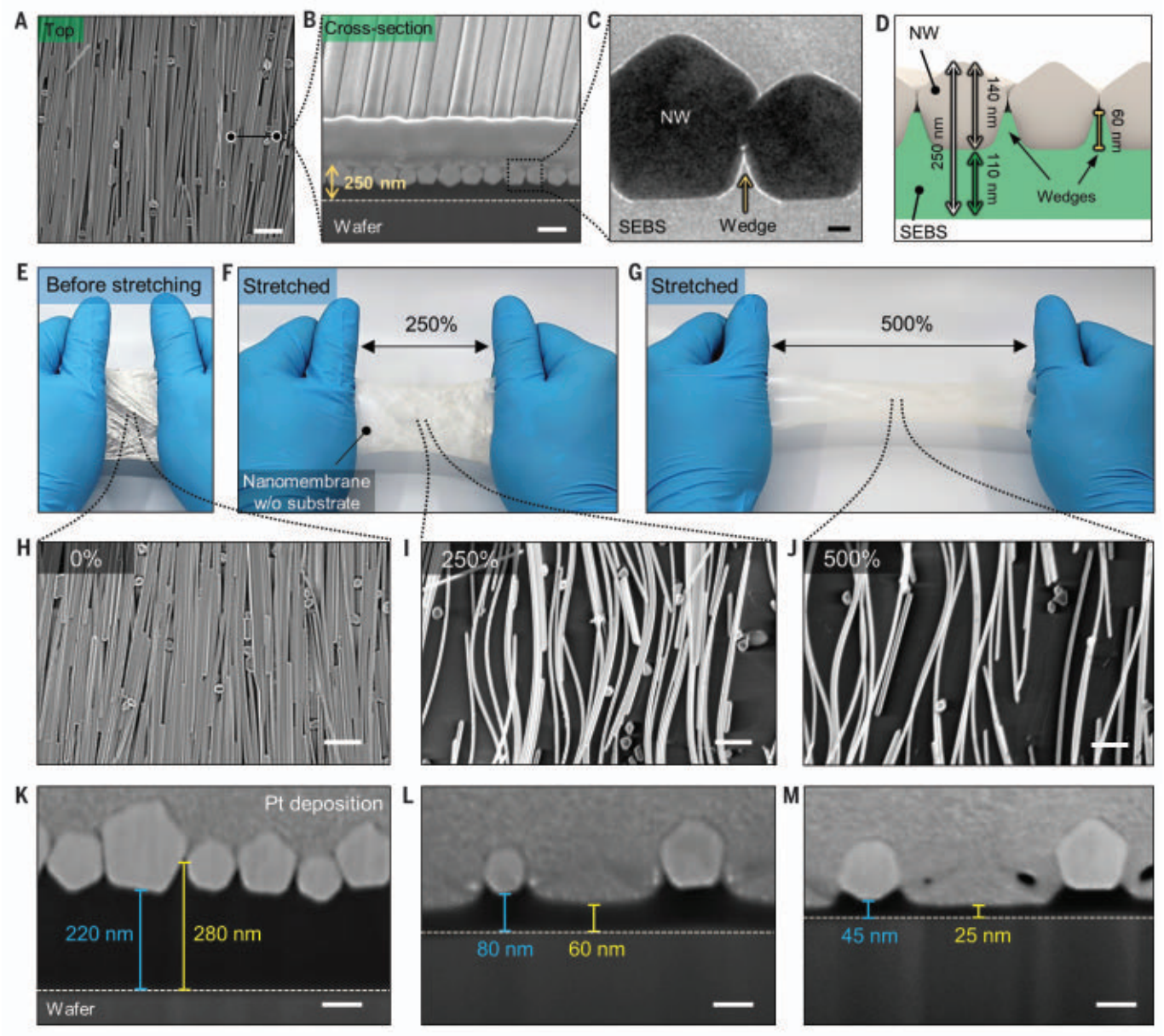


Fig. 2. Structural and mechanical characterization of the nanomembrane. (**A** to **C**) Microscopy images that show the structure of an ultrathin nanomembrane. (A) SEM image (top view) of closely packed NWs in the nanomembrane (scale bar, 2 μm). (B) Magnified SEM image of the cross section of the nanomembrane (scale bar, 200 nm). (C) Higher-magnification TEM image (cross-sectional view) of the assembled NWs fixed in the ultrathin elastomer layer. The elastomer wedge in between NWs is marked by a yellow arrow (scale bar, 20 nm). (**D**) A schematic illustration of the nanomembrane cross section that describes its detailed structure and dimensions. (**E** to **G**) Photographs that show the high elasticity of a free-standing nanomembrane for an original state (E) and stretched states under 250% strain (F) and 500% strain (G). (**H** to **J**) SEM images (top view) that show each state of the nanomembrane (scale bar, 2 μm). (**K** to **M**) SEM images (cross-sectional view) that show each state of the nanomembrane in Fig. 2, H to J (scale bar, 100 nm). The elastomer thickness in between NWs (yellow) and that under the NW (sky blue) are indicated. The dotted line in cross-sectional SEM images shows the interface between SEBS and a handling wafer. Platinum (Pt) is deposited on top to protect NWs during ion milling.

embedded in gums. Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) images for top and cross-sectional views of the nanomembrane (Fig. 2, A to C) confirm the structure. Detailed dimensions are described in a schematic illustration (Fig. 2D). The nanomembrane, with a thickness of ~250 nm, consists of a monolayer of NWs with a diameter of ~140 nm and an elastomer layer with a thickness of ~110 nm. Periodic elastomer wedges with a height of ~60 nm in between the NWs support the structure. The nanomembrane thickness depends on the elastomer thickness, which is controlled by the amount of the elastomer in the nanocomposite solution (fig. S5, A to C).

This gumlike elastomer structure efficiently dissipates the induced strain. A typical nanocomposite in which rigid metallic fillers are fully embedded inside the elastomer matrix exhibits a high level of stress focused on the interface between filler and elastomer under mechanical deformations (fig. S5, D to F). However, the gumlike structure in which NWs are

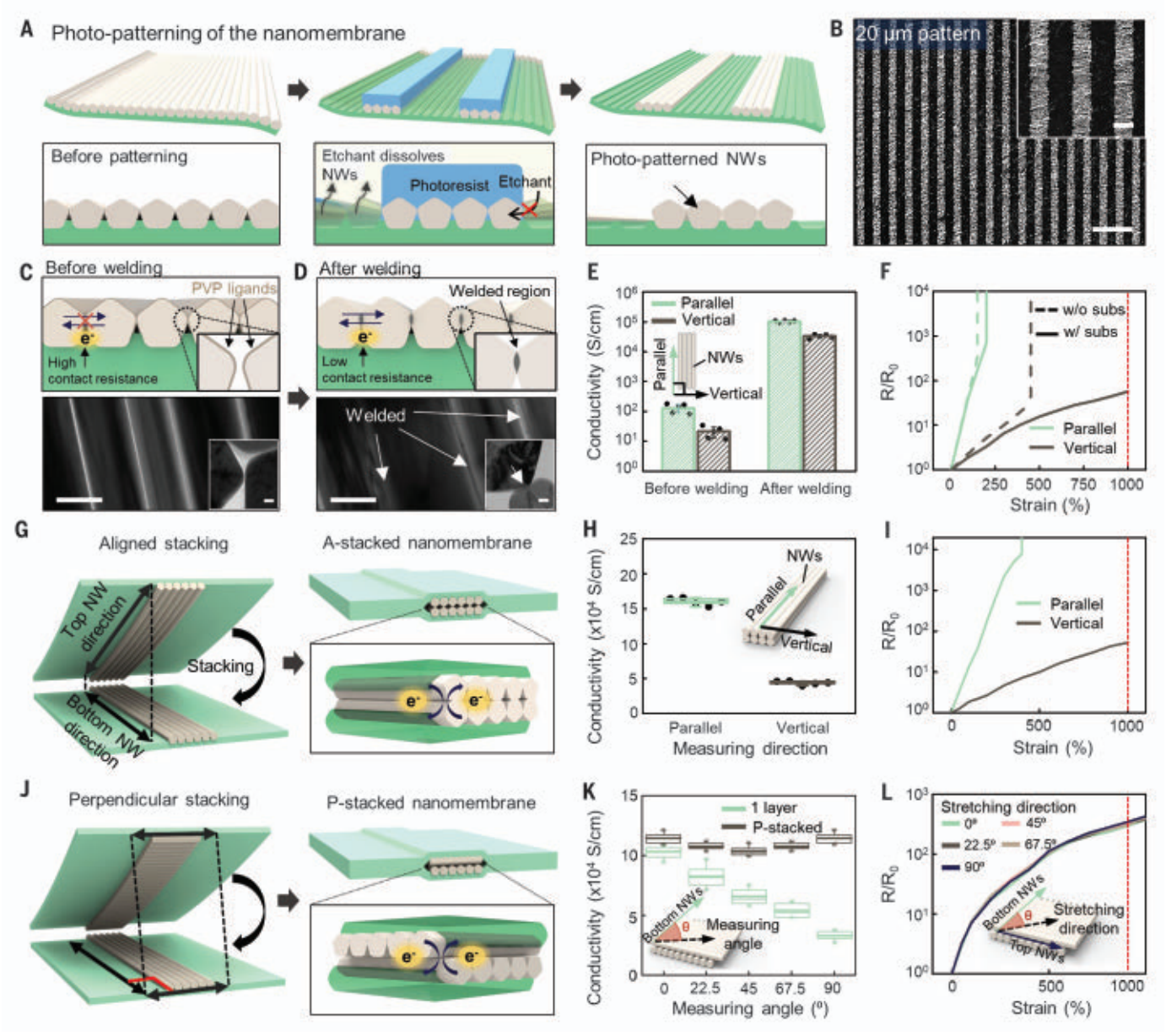


Fig. 3. Electrical performance of the patterned nanomembrane after cold welding and stacking. (**A** and **B**) High-resolution patterning of the nanomembrane using photolithography. (A) Schematic illustrations of the patterning process. (B) SEM image of a nanomembrane with a 20-μm line-and-space pattern (scale bar, 100 μm). An inset shows a magnified view (scale bar, 20 μm). (**C** and **D**) Schematic illustrations (cross-sectional view) and TEM images [top view and cross-sectional view (inset)] (C) before and (D) after cold welding (scale bar, 100 nm; inset scale bar, 20 nm). (**E**) Conductivity enhancement after cold welding. Parallel and Vertical indicate the measurement direction against the direction of NWs (n = 5, mean ± SD). (**F**) Normalized resistance under applied strains. Solid lines are for a nanomembrane on an elastic substrate, and dashed lines are for a free-standing nanomembrane without a substrate. A red dashed line indicates the point at which mechanical failure of the substrate occurs. (**G** to **I**) Performance improvement after aligned stacking of two nanomembranes. (G) Schematic illustrations of two stacked nanomembranes. (H) Conductivity of the stacked nanomembrane, depending on measurement directions (n = 5). (I) Normalized resistance under applied strains for two stretching directions on an elastic substrate. (**J** to **L**) Removal of directionality in conductance after perpendicular stacking. (J) Schematic illustrations of two stacked nanomembranes. (K) Conductivity of the stacked nanomembrane compared with conductivity of a single-layer nanomembrane, depending on measurement directions (n = 5). (L) Normalized resistance under applied strains measured in various stretching directions on an elastic substrate.

partially embedded shows much less stress at the interface under the same mechanical deformations (fig. S5, G to I). This results in outstanding elasticity even with a high weight fraction of NWs (>80 wt %). The maximum elongation is 540% (Fig. 2, E to G), which is almost comparable to that of the bare elastomer membrane (570%) (fig. S5J).

The applied strain is dissipated mostly by the elastomer layer, in particular by wedge regions, as confirmed by SEM during stretching of the nanomembrane (Fig. 2, H to M). Before the stretching, the elastomer thickness in between NWs (~280 nm; yellow) is thicker than that under NWs (~220 nm; sky blue) (Fig. 2K).

However, with strains of 250 and 500%, the elastomer thickness in between NWs becomes ~60 and ~25 nm (yellow), and that under NWs becomes ~80 and ~45 nm (sky blue), respectively (Fig. 2, L and M). The applied strain is mostly dissipated by the elastomer in between NWs (i.e., wedge region), which minimizes the induced strain at the interface between NWs and elastomer, leading to a large elongation of the nanomembrane.

The teethlike NW structure allows high-resolution patterning of NWs using photolithography (Fig. 3A). NWs inside the desired pattern are protected by the photoresist, and NWs outside the pattern can be etched. Because NWs are partially exposed from the elastomer, their etching is facile (Fig. 3B). Various patterns can be fabricated, and the entire patterns are stretchable (fig. S6, A and B). The compactly assembled NWs ensure connectivity between NWs even after the high-resolution patterning (fig. S6C).

Contacts between NWs can be reinforced further by cold welding of the partially exposed NWs. The surface of NWs is covered by PVP ligands, which cause a slight separation between NWs, according to TEM (Fig. 3C). When treated with salty water such as a sodium chloride solution, PVP can be removed from the NW surface (*26*). Afterward, the capillary force generated during water evaporation induces cold welding of NWs (fig. S7, A to D) (*27*, *28*). Firm connections across NWs are made (Fig. 3D), which enhance conductivity significantly (Fig. 3E). Conductivity can differ depending on measurement directions, 103,100 or 32,900 S/cm, parallel or vertical to the direction of NWs, respectively.

The nanomembrane remains highly conductive up to a strain of 200% in the parallel direction and >1000% in the vertical direction (Fig. 3F). For the stretching test, the nanomembrane was mounted on an elastic substrate, because stretchable electrodes are often mounted on elastic substrates. Without the elastic substrate, the strain could not be applied uniformly during the stretching test, leading to a lower stretchability of 150% (parallel) and 450% (vertical). The elastic substrate helps the uniform distribution of the applied strain.

Without welding, NWs are disconnected under stretching (fig. S7, E to G). However, the NWs connected by the cold welding maintain the electrical connection under the stretching (fig. S7, H to J). The degree of welding differs depending on the distance between NWs. Although NWs are compactly assembled in the nanomembrane, slight differences in the spacing between NWs exist. This small difference in the spacing can result in the large difference in the capillary force between the adjacent NWs, because the capillary force is inversely proportional to the spacing (*29*). Thus, welding occurs in the relatively small-gap regions only. Therefore, during stretching, the welded regions maintain connection, but the nonwelded regions separate. If sintering occurs instead of local welding, as when thin NWs are used instead of thick NWs, the NWs become a brittle silver film (fig. S8, A to C). In such a case, the capillary force between NWs can be reduced by using a solvent with a relatively low surface tension such as isopropyl alcohol to induce local welding rather than sintering (fig. S8D).

Conductivity can be maximized (Fig. 3, G to I) or asymmetry of conductance can be improved (Fig. 3, J to L) by stacking two nanomembrane layers (fig. S9). When two nanomembranes are stacked with the NWs aligned to each other (Fig. 3G), a maximum conductivity of 165,700 S/cm is achieved (Fig. 3H). The stacked nanomembrane remains conductive up to ~400% strain in the parallel direction and >1000% in the vertical direction (Fig. 3I). When two nanomembranes are stacked with the NWs perpendicular to each other (Fig. 3J), a conductivity of >100,000 S/cm is achieved regardless of measurement directions (Fig. 3K). Furthermore, the stacked nanomembrane remains conductive even under 1000% strain, regardless of stretching directions (Fig. 3L).

Nanomembranes that use other metal nanomaterials, such as Ag-Au NWs, Ag NPs, and Au NPs, can be also fabricated (fig. S10A). The nanomembrane with Ag-Au NWs exhibits similar properties and performance to that with Ag NWs. Meanwhile, nanomembranes with Ag NPs or Au NPs have different features from those with NWs, such as symmetrical conductance, lower conductivity, and higher sensitivity to external strains. After the cyclic stretching test, the nanomembranes fabricated with Ag NWs, Ag-Au core-shell NWs, Ag NPs, and Au NPs still showed stable conducting performance (fig. S11). Nanomembranes can be fabricated by using other elastomers such as TPU and SIS (fig. S10B).

Photopatterned nanomembranes can be applied to skin electronics (supplementary text, section S1). Devices ranging from a simple skin-mounted electrode (fig. S12) to a multifunctional sensor array composed of electrophysiology, temperature, strain, and humidity sensors can be fabricated by integrating multiple layers of patterned nanomembranes (figs. S13 and S14). Each sensor array is connected to interconnections by vertical interconnect accesses (figs. S15 and S16). The multifunctional sensor array exhibits reliable operation on human skin.

We have presented a method to fabricate highly conductive, elastic, and ultrathin nanomembranes (fig. S17). Exceptional material properties are achieved by the structure of the nanomembrane, which is a monolayer of assembled nanomaterials partially embedded in an ultrathin elastomer membrane (supplementary text, section S2). Photolithography can be used for high-resolution patterning and stacking strategies, and welding processes enhance the performance of the nanomembrane. By using patterned nanomembranes, a multifunctional skin-mounted sensor array can be fabricated.

REFERENCES AND NOTES

1. D.-H. Kim *et al.*, *Science* **333**, 838–843 (2011).
2. Z. Bao, X. Chen, *Adv. Mater.* **28**, 4177–4179 (2016).
3. D. Son *et al.*, *Nat. Nanotechnol.* **13**, 1057–1065 (2018).
4. S. Gong *et al.*, *Nat. Commun.* **5**, 3132 (2014).
5. I. You *et al.*, *Science* **370**, 961–965 (2020).
6. S. Gong *et al.*, *Adv. Mater.* **31**, e1903789 (2019).
7. K. Wang *et al.*, *Adv. Funct. Mater.* **2021**, 2008347 (2021).
8. S. Kang *et al.*, *Sci. Adv.* **4**, eaas8772 (2018).
9. P. Grandgeorge *et al.*, *Science* **360**, 296–299 (2018).
10. Z. Liu *et al.*, *Adv. Mater.* **30**, 1704229 (2018).
11. T. Someya, Z. Bao, G. G. Malliaras, *Nature* **540**, 379–385 (2016).
12. Z. Jiang *et al.*, *Adv. Mater.* **31**, e1903446 (2019).
13. T. C. Shyu *et al.*, *Nat. Mater.* **14**, 785–789 (2015).
14. Y. Kim *et al.*, *Nature* **500**, 59–63 (2013).
15. N. Matsuhisa *et al.*, *Nat. Mater.* **16**, 834–840 (2017).
16. Z. Ma *et al.*, *Nat. Mater.* **20**, 859–868 (2021).
17. S. Veerapandian *et al.*, *Nat. Mater.* **20**, 533–540 (2021).
18. S. Lee *et al.*, *Science* **370**, 966–970 (2020).
19. J. Xu *et al.*, *Science* **355**, 59–64 (2017).
20. A. Miyamoto *et al.*, *Nat. Nanotechnol.* **12**, 907–913 (2017).
21. S. Wang *et al.*, *Nature* **555**, 83–88 (2018).
22. J. Lyu *et al.*, *Adv. Funct. Mater.* **26**, 8435–8445 (2016).
23. S. Choi *et al.*, *Nat. Nanotechnol.* **13**, 1048–1056 (2018).
24. N. Matsuhisa *et al.*, *Nat. Commun.* **6**, 7461 (2015).
25. Y. Lu *et al.*, *ACS Appl. Mater. Interfaces* **10**, 2093–2104 (2018).
26. M. Grouchko, A. Kamyshny, C. F. Mihailescu, D. F. Anghel, S. Magdassi, *ACS Nano* **5**, 3354–3359 (2011).
27. Y. Liu *et al.*, *Nano Lett.* **17**, 1090–1096 (2017).
28. X. Li *et al.*, *Nat. Commun.* **10**, 3514 (2019).
29. Y. I. Rabinovich, M. S. Esayanur, B. M. Moudgil, *Langmuir* **21**, 10992–10997 (2005).

ACKNOWLEDGMENTS

Funding: This research was supported by the Institute for Basic Science (IBS-R006-D1 and IBS-R006-A1). This research was also supported by the National Research Foundation (NRF) of Korea (2017M3D1A1039288/2018R1A4A1025623). **Author contributions:** D.J., C.L., H.J.S., T.H., and D.-H.K. conceived the ideas, designed the experiments, and wrote the manuscript. D.J., C.L., H.J.S., Y.K., C.P., S.I.H., S.-H.S., K.W.C., G.D.C., D.C.K., J.H.Koo., T.H., and D.-H.K. performed experiments and analysis. J.J. and J.H.Kim. performed computer simulation. **Competing interests:** D.-H.K., T.H., D.J., C.L., and H.J.S. are inventers on a patent application related to this work (KR 10-2021-0072433, 3 June 2021). The authors declare no other competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/1022/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S26
Movies S1 to S5
References (*30*–*53*)

11 March 2021; accepted 23 July 2021
10.1126/science.abh4357

PALEOECOLOGY

A positive relationship between functional redundancy and temperature in Cenozoic marine ecosystems

T. M. Womack[1]*, J. S. Crampton[1], M. J. Hannah[1], K. S. Collins[2]

The long-term effects of climate change on biodiversity and biogeographic patterns are uncertain. There are known relationships between geographic area and both the number of species and the number of ecological functional groups—termed the species-area relationship and the functional diversity–area relationship, respectively. We show that there is a positive relationship between the number of species in an area, the number of ecological functional groups, and oceanic temperature in the shallow-marine fossil record of New Zealand over a time span of ~40 million years. One implication of this relationship is that functional redundancy increases with temperature. This reveals a long-lived and persistent association between the spatial structuring of biodiversity, the temperature-dependence of functional redundancy, and shallow-marine biodiversity in mid-latitudes.

One of the emerging objectives of paleobiology is to determine drivers of spatial and temporal distributions of biodiversity to assess future ecological impacts of climatic warming. Understanding the spatial structuring of biodiversity is fundamental to achieving this objective. Biodiversity in the fossil record can be measured in terms of species richness and functional richness—the latter measured as the number of unique functional groups (defined on the basis of common functional traits shared between species), here used as a proxy for ecological niche (*1*). The number of species occupying a geographic area is proportional to the size of that area, a pattern referred to as the species-area relationship. This has been documented in the fossil record (*2*) and conforms to the well-established power function model observed in modern ecology ($S = cA^z$, where S is species richness, c and z are constants, and A is area) (*3*). The classical form of the species-area relationship has been widely used to assess the impact of anthropogenic climate change and habitat loss on modern ecosystems (*4*, *5*). On the other hand, spatial scaling of functional richness—herein referred to as the functional diversity–area relationship—has not been quantified in the fossil record and is understudied in modern ecology. Notably, it is not known whether the relationships between species and functional richness at increasing spatial scales are conserved through time or whether the two respond differently to factors that might influence the geographic distribution of modern organisms in response to future climate change.

Recent research across the modern marine latitudinal diversity gradient records high levels of functional redundancy—the duplication of functional roles by multiple species—in tropical-subtropical latitudes that decrease in temperate latitudes (*6*). This pattern suggests that functional redundancy may be driven by latitudinal gradients of climate and temperature and, in the case of marine molluscs, is likely controlled by total resource abundance (*6*). A similar pattern of functional redundancy relative to species richness has also been observed at regional scales in modern mollusc ecosystems (*7*). If functional redundancy is mediated by climate, and in particular temperature, we would anticipate that the spatial relationship between species and functional richness would vary through time and be correlated with climate proxies over geological time scales, assuming these relationships reach an ecological equilibrium. This is important because evidence from modern ecology suggests that functional redundancy may increase ecosystem resilience to future environmental change (*8*).

We quantify and compare the species-area relationship and the functional diversity–area relationship using the Cenozoic (fig. S1) fossil record of New Zealand shallow-marine molluscs, and we assess their relationships to regional climate—measured as oceanic temperature—and several other potentially important explanatory variables, including climatic variability, the quality of the geological rock record, and sampling effort (*1*). New Zealand has remained biogeographically isolated for the duration of the Cenozoic with a high degree of endemism, and speciation is predominately attributed to in situ processes (*9*). Thus, historic changes in New Zealand's biogeography are unlikely to be related to the wholesale migration of species. Furthermore, oceanic temperature is considered to be one important driver of marine species diversity through space and time (*10*–*12*), and current climate models project substantial changes in both surface and deep ocean temperatures over the coming century (*12*).

We explicitly standardize our data for the uneven distribution of fossil collections temporally (*13*) and spatially using summed minimum spanning tree (MST) length as our measure of geographic scale (*14*)—a measure of the minimum summed length of linear segments that connects all coordinates within a sample (*1*). For simplicity, we refer to the relationship between the species-area relationship and the functional diversity–area relationship as the functional redundancy relationship. This describes, for a given time bin, the relationship between values of the species-area relationship (on the abscissa) and the functional diversity–area relationship at discrete steps of summed MST length (*1*). The slope of the functional redundancy relationship reflects an element of functional redundancy, where shallower slopes indicate an increase in functional redundancy.

The standardized species-area relationship and functional diversity–area relationship in the Cenozoic shallow-marine molluscan fossil record of New Zealand conform to the power function model observed in modern ecological studies (*3*) (figs. S2 and S3 and table S1). Two time bins have inadequate sampling and spatial coverage (bins 2 and 3) (figs. S2 and S3) and are omitted from further analyses, although their inclusion or exclusion does not change the main conclusions of this study (*1*). The functional redundancy relationship can also be described using a power function model (table S1 and fig. S4). The slopes of the functional redundancy relationship correlate with oceanic temperature in the Pacific Ocean (Spearman rank-order correlation coefficient, $r_s = -0.829$; $P = 0.003$) (see Fig. 1 for a graphical workflow and Fig. 2), where lower slope values correlate with higher oceanic temperatures. We account for serial autocorrelation using both the Spearman rank-order correlation coefficient with phase-randomized surrogate data (table S2) and generalized least squares (GLS) autoregressive (AR1) models (table S3) (*1*). Additionally, there is evidence that species richness, adjusted for uneven spatial and temporal sampling, is positively related to the combined effect of oceanic temperature and climatic variability (table S3). However, we acknowledge that ocean temperature and climate variability are intrinsically linked, and dynamics of climate variability during the Miocene climatic optimum [~17 to 14 million years ago (Ma)] and the Plio-Pleistocene "Icehouse" were very different (*15*); with existing data, we cannot fully separate the effects of temperature and climate variability on New Zealand mollusc species richness. Lastly, we

[1]School of Geography, Environment and Earth Sciences, Victoria University of Wellington, Wellington, New Zealand. [2]Department of Earth Sciences, The Natural History Museum, London SW7 5BD, UK.
*Corresponding author. Email: tom.womack@vuw.ac.nz

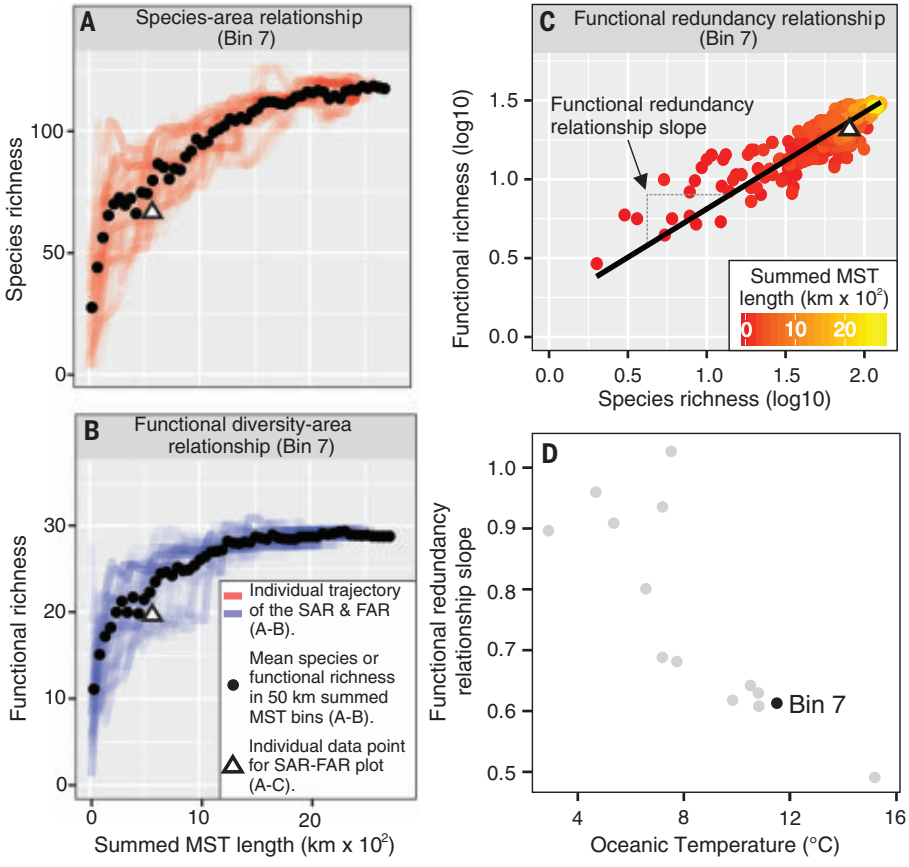


Fig. 1. Graphical workflow illustrating the derivation of the functional redundancy relationship. (**A** and **B**) The species-area relationship (SAR) (A) and the functional diversity–area relationship (FAR) (B) were plotted against the summed MST length for each time bin (using bin 7 as an example). The red and blue lines mark individual spatial trajectories of the species-area relationship and the functional diversity–area relationship, respectively. (**C**) We plotted the individual data points of species (A) and functional richness (B) at corresponding steps of spatial scale from the individual trajectories of the species-area relationship and the functional diversity–area relationship (referred to as the SAR-FAR plot in the key) (individual points are color coded to the corresponding summed MST length), as illustrated using a single point (triangular symbol) in (A) to (C). We measured the reduced major axis (RMA) slope of the functional redundancy relationship for each time bin (slope for bin 7 is highlighted). The steepness of this slope reflects an element of functional redundancy, where shallower slopes indicate an increase in functional redundancy. (**D**) We plotted the slopes for each time bin against oceanic temperature (same as Fig. 2B). The variables represented in (D) are not independent; this is corrected for within our analyses (*1*).

find no strong or statistically significant evidence that the primary results discussed here result from temporal variability in sampling effort, time bin duration, or the quality of the geological rock record (tables S2 and S3).

Our main conclusion is that at warmer oceanic temperatures, species richness accumulates faster than functional richness with increasing geographic scale. At cooler oceanic temperatures, species richness accumulates at a slower rate. This suggests that functional redundancy is relatively elevated at warmer oceanic temperatures—i.e., under these conditions, on average, more species occupy each functional group (niche). Standardized measures of functional evenness—a measure of how uniformly species are distributed amongst functional groups—suggest that species are generally less-evenly distributed between functional groups during the global warmth of the Miocene climatic optimum (fig. S6), where functional evenness is low. Functional evenness subsequently increases with global and regional cooling in New Zealand during the middle Miocene climate transition. A similar pattern is present in the modern latitudinal gradient, where bivalve species and genera are less-evenly spread among functional groups at low latitudes and become more-evenly dispersed poleward (*6*). The decrease in functional evenness during the Miocene climatic optimum coincides with a substantial decrease in the rate and magnitude of temporal turnover of functional groups—the change within and between functional groups over time (*1*). This decrease in temporal turnover is facilitated predominately by a decrease in the replacement of functional groups (fig. S6), which indicates that the increase in species richness during the Miocene climatic optimum is driven largely by expansion within existing functional groups (i.e., existing ecological niches) rather than the addition of new groups. At a representative fixed spatial scale, functional richness remains relatively static through much of the Cenozoic, except for the Plio-Pleistocene (5.33 to 0 Ma) (Fig. 2 and fig. S3). Together, these findings suggest that increases in regional molluscan species richness during the warm Miocene climatic optimum are primarily a result of increased niche packing (i.e., more species occupying the same niche rather than niche expansion), and the subsequent decrease in species richness during the Miocene climatic transition has minimal impact on functional richness relative to species richness.

The correlation between temperature and functional redundancy implies a relationship between oceanic temperature and standing diversity. This is further supported by the positive relationship between species richness, oceanic temperature, and climatic variability. Whereas the notion of an ecological carrying capacity (a limit on biodiversity) is debated (*16*–*18*), our results suggest that lower oceanic temperatures impose an ecological constraint on species richness in New Zealand, particularly before the Plio-Pleistocene (>5.33 Ma). When we control for geographic scale, we find slightly elevated species richness and elevated functional richness during three Plio-Pleistocene time bins (bins 13, 15, and 16; see Fig. 2). The Plio-Pleistocene is a period of known climatic variability (*15*) and, within New Zealand, increased landscape complexity related to accelerated tectonic activity (*19*), maturation of oceanic fronts, and intensification of marine latitudinal temperature gradients (*20*), and it is initially marked by a period of transient warming (*21*). We find evidence that oceanic temperature and climatic variability contribute positively to species richness, which explains, in part, the early Pliocene and Pleistocene increase in species richness. By contrast, we find no consistent explanation for increases in functional richness. However, we do observe a substantial increase in temporal turnover of functional groups at the Plio-Pleistocene boundary, facilitated predominately by an increase in the replacement

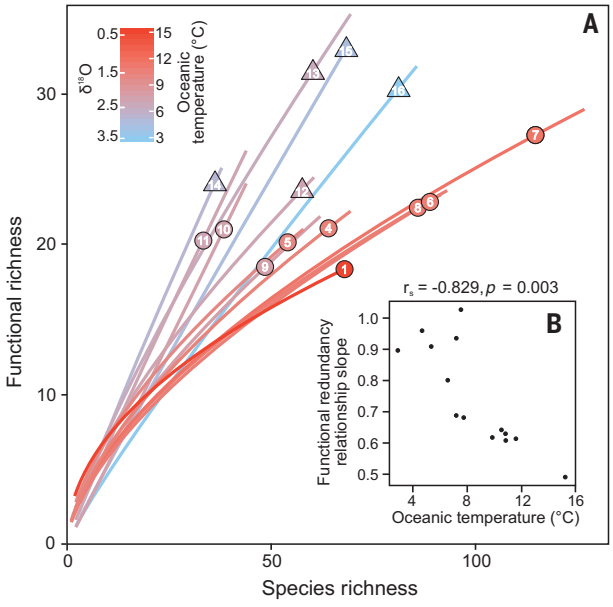


Fig. 2. Relationship between functional redundancy and oceanic temperature, 45.7 to 0 Ma. (**A**) Functional redundancy relationship plotted in linear space for each time bin (cf. Fig. 1C, black regression line, but plotted in linear space). Individual trajectories of the functional redundancy relationship are color coded to $\delta^{18}O$ and oceanic temperature (in degrees Celsius). Points are labeled by time bin (fig. S1), ordered from oldest (bin 1) to youngest (bin 16), and represent values of species and functional richness (*1*) at a representative spatial scale (summed MST length, 1300 to 1350 km). Shallower trajectories of individual lines represent higher functional redundancy. Triangular points represent Plio-Pleistocene time bins, and circular points represent time bins older than the Pliocene (>5.33 Ma). (**B**) Scatterplot of the slope of the functional redundancy relationship and oceanic temperature (in degrees Celsius) in each time bin. The variables represented in (B) are not independent; this is corrected for within our analyses (*1*).

of functional groups (fig. S6). Although it is not possible to determine the exact causes of the increase in functional richness during the Plio-Pleistocene, one plausible cause for the increase in functional group turnover at the Pliocene boundary, and perhaps functional richness, may be the intensification of marine latitudinal temperature gradients around New Zealand. Whereas oceanic temperature is implicated here, the exact mechanisms responsible for the temperature–functional redundancy relationship are not clear and could feasibly be related to both abiotic and biotic factors. Overall, our findings suggest that New Zealand's species richness fluctuates with oceanic temperature, corroborating recent research that suggests macroevolutionary models should incorporate time-varying (i.e., not static) carrying capacities of species richness (*18*).

Predicting the ecological response to climate change requires an understanding of how ecosystems have responded to climatic fluctuations in the past, within a spatial context. This is particularly important for temperate marine ecosystems, like those of New Zealand, which span climatic, oceanographic, and biogeochemical boundaries (*22*), where the largest changes in species and functional richness are observed (*6*). We find that despite key differences in both temporal and spatial resolution between paleontological and ecological data, the functional diversity–area relationship in the fossil record conforms to similar patterns observed in modern ecological studies (*23*) and, notably, is consistent over long spans of time. This suggests that the functional diversity–area relationship is as persistent and predictable as the species-area relationship. Furthermore, the observed correlation between functional redundancy and oceanic temperature over the last ~40 million years is consistent with observations from the modern latitudinal diversity gradient (*6*), suggesting a long-lived relationship of regional and global importance. Taken at face value, our results suggest that oceanic temperature should increase net species richness and functional redundancy in New Zealand over long time spans (i.e., multicentennial to millennial), particularly as we shift to a climate more representative of pre-Pleistocene conditions (*24*). This provides a baseline for what should be expected—at ecological equilibrium—from natural warming in regional, temperate shallow-marine ecosystems.

REFERENCES AND NOTES

1. Materials and methods are available as supplementary materials online.
2. A. D. Barnosky, M. A. Carrasco, E. B. Davis, *PLOS Biol.* **3**, e266 (2005).
3. M. L. Rosenzweig, *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
4. J. M. Halley, V. Sgardeli, K. A. Triantis, *Glob. Ecol. Biogeogr.* **23**, 113–123 (2014).
5. O. T. Lewis, *Phil. Trans. R. Soc. B* **361**, 163–171 (2006).
6. M. Schumm *et al.*, *Proc. R. Soc. B.* **286**, 20190745 (2019).
7. K. S. Collins, S. M. Edie, T. Gao, R. Bieler, D. Jablonski, *PLOS ONE* **14**, e0221490 (2019).
8. C. R. Biggs *et al.*, *Ecosphere* **11**, e03184 (2020).
9. J. S. Crampton *et al.*, *Paleobiology* **32**, 509–532 (2006).
10. L. H. Antão *et al.*, *Nat. Ecol. Evol.* **4**, 927–933 (2020).
11. D. H. Erwin, *Curr. Biol.* **19**, R575–R583 (2009).
12. N. L. Bindoff *et al.*, in *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner *et al.*, Eds. (Intergovernmental Panel on Climate Change, 2019), chap. 5.
13. J. Alroy, *Paleobiology* **46**, 158–175 (2020).
14. T. M. Womack, J. S. Crampton, M. J. Hannah, *Paleobiology* **47**, 39–53 (2021).
15. T. Westerhold *et al.*, *Science* **369**, 1383–1387 (2020).
16. D. L. Rabosky, A. H. Hurlbert, *Am. Nat.* **185**, 572–583 (2015).
17. L. J. Harmon, S. Harrison, *Am. Nat.* **185**, 584–593 (2015).
18. C. R. Marshall, T. B. Quental, *Phil. Trans. R. Soc. B* **371**, 20150217 (2016).
19. P. R. King, T. R. Naish, G. H. Browne, B. D. Field, S. W. Edbrooke, *Cretaceous to Recent Sedimentary Patterns in New Zealand* (Institute of Geological and Nuclear Sciences, folio series 1, 1999).
20. C. S. Nelson, P. J. Cooke, *N. Z. J. Geol. Geophys.* **44**, 535–553 (2001).
21. T. D. Herbert *et al.*, *Nat. Geosci.* **9**, 843–847 (2016).
22. G. Reygondeau *et al.*, *Global Biogeochem. Cycles* **27**, 1046–1058 (2013).
23. R. J. Whittaker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13709–13714 (2014).
24. K. D. Burke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 13288–13293 (2018).

ACKNOWLEDGMENTS

We acknowledge use of the data in the New Zealand Fossil Record File, administered by the Geological Society of New Zealand and GNS Sciences (www.fred.org.nz). We thank J. Alroy for valuable discussions on estimating area in the fossil record and estimating species and functional richness using different standardization methods, and we thank three anonymous reviewers for their insightful comments, which greatly improved the manuscript. **Funding:** This work is funded by Victoria University of Wellington through the Commonwealth Scholarship to T.M.W. **Author contributions:** T.M.W. designed the research, wrote the manuscript, carried out the analyses, and prepared the figures, under the guidance of J.S.C. and M.J.H. K.S.C. updated the ecospace compiled from the synoptic dataset by T.M.W. and contributed to interpretation of results in relation to the modern latitudinal diversity gradient. All authors discussed the results and reviewed the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Occurrence data are available for download from the New Zealand Fossil Record Electronic Database (FRED) (https://fred.org.nz/). The complete R code along with the relevant data files are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/1027/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S4
References (*25*–*54*)
MDAR Reproducibility Checklist
Data S1

29 November 2020; resubmitted 23 February 2021
Accepted 28 July 2021
10.1126/science.abf8732

HUMAN GENETICS

Population sequencing data reveal a compendium of mutational processes in the human germ line

Vladimir B. Seplyarskiy[1,2]†, Ruslan A. Soldatov[2]†, Evan Koch[1,2], Ryan J. McGinty[1,2], Jakob M. Goldmann[3], Ryan D. Hernandez[4,5], Kathleen Barnes[6], Adolfo Correa[7,8,9], Esteban G. Burchard[5,10], Patrick T. Ellinor[11], Stephen T. McGarvey[12,13,14], Braxton D. Mitchell[15,16,17], Ramachandran S. Vasan[18,19], Susan Redline[20,21], Edwin Silverman[22], Scott T. Weiss[20,21,22], Donna K. Arnett[23], John Blangero[24,25], Eric Boerwinkle[26,27], Jiang He[28,29], Courtney Montgomery[30], D. C. Rao[31], Jerome I. Rotter[32], Kent D. Taylor[32], Jennifer A. Brody[33], Yii-Der Ida Chen[34], Lisa de las Fuentes[31,35], Chii-Min Hwu[36], Stephen S. Rich[37], Ani W. Manichaikul[37], Josyf C. Mychaleckyj[37], Nicholette D. Palmer[38], Jennifer A. Smith[39,40], Sharon L. R. Kardia[40], Patricia A. Peyser[40], Lawrence F. Bielak[40], Timothy D. O'Connor[41,42,43], Leslie S. Emery[44], NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium‡, TOPMed Population Genetics Working Group, Christian Gilissen[3], Wendy S. W. Wong[45], Peter V. Kharchenko[2], Shamil Sunyaev[1,2]*

Biological mechanisms underlying human germline mutations remain largely unknown. We statistically decompose variation in the rate and spectra of mutations along the genome using volume-regularized nonnegative matrix factorization. The analysis of a sequencing dataset (TOPMed) reveals nine processes that explain the variation in mutation properties between loci. We provide a biological interpretation for seven of these processes. We associate one process with bulky DNA lesions that are resolved asymmetrically with respect to transcription and replication. Two processes track direction of replication fork and replication timing, respectively. We identify a mutagenic effect of active demethylation primarily acting in regulatory regions and a mutagenic effect of long interspersed nuclear elements. We localize a mutagenic process specific to oocytes from population sequencing data. This process appears transcriptionally asymmetric.

Decades of experimental research in mutagenesis revealed the various error modes of DNA replication and repair (*1*, *2*) but did not elucidate which mechanisms are primarily responsible for human germline mutations. Statistical analysis of sequencing datasets can quantify contributions of relevant mechanisms.

Cancer genomics has been propelled by the analysis of "mutation signatures" (*3*). Signature extraction relies on differential mutagen exposure of tumor samples and is not transferable to germline mutation, beyond insights from comparing human populations (*4*).

Here, we use variation of mutation rate along the genome to model germline mutagenesis. Our model assumes that several mechanisms generate mutations. We approximate these mechanisms by processes characterized by spectra of 192 trinucleotide mutation types and variable intensities along the genome (Fig. 1A). Inference of mutational processes from variability of mutational spectra across genomic loci represents a classic nonnegative matrix factorization (NMF) problem.

NMF separates a set of nonnegative source signals (here, mutational processes) from observed nonnegative signal mixtures (here, mutation frequencies). However, NMF can have many solutions with the same quality of approximation (*5*). To find identifiable solutions, applications in cancer assume that tumors are exposed to a few mutagenic forces characterized by specific mutation types (*6*).

Classic NMF performs poorly in our case (fig. S1K), likely because mutagenic forces with complex spectra act in most loci. Recently developed "volume-regularized" NMF (vrnmf) (*6*, *7*) guarantees a unique solution under mild assumptions of sufficiently spread source signals. Vrnmf finds the most distinct (positively defined) mutational spectra (Fig. 1B and fig. S1A). For germline mutations, our implementation of vrnmf (*8*) delivers a unique interpretable solution, outperforming standard NMF in simulations and data (fig. S1, K and C).

A powerful way to assess the biological relevance of the inferred processes is provided by the symmetry between antiparallel strands of DNA. Strand-specific footprints of molecular machineries such as transcription and replication break this symmetry. Mutational mechanisms coupled with these machineries are strand dependent. For example, A>G mutations are depleted within genes on the transcribed strand owing to the action of transcription-coupled repair (TCR) (*2*). Mutation processes uncoupled from the action of strand-specific machineries are strand independent (Fig. 1C).

We assign mutation types with respect to the genome reference independently of the direction of transcription and replication. For some genes, the reference strand is transcribed,

[1]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [2]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [3]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, Netherlands. [4]Quantitative Life Sciences, McGill University, Montreal, QC, Canada. [5]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA. [6]Department of Medicine, University of Colorado Denver, Aurora, CO 80045, USA. [7]Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [8]Department of Pediatrics, University of Mississippi Medical Center, Jackson, MS, USA. [9]Department of Population Health Science, University of Mississippi Medical Center, Jackson, MS, USA. [10]Department of Medicine, University of California, San Francisco, CA, USA. [11]Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [12]International Health Institute, Brown University, Providence, RI, USA. [13]Department of Epidemiology, Brown University, Providence, RI, USA. [14]Department of Anthropology, Brown University, Providence, RI, USA. [15]Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [16]Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [17]Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA. [18]Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [19]Framingham Heart Study, Framingham, MA, USA. [20]Department of Medicine, Harvard Medical School, Boston, MA, USA. [21]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [22]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [23]Department of Epidemiology, University of Kentucky, Lexington, KY, USA. [24]Department of Human Genetics, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA. [25]South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA. [26]University of Texas Health Science Center at Houston, Houston, TX, USA. [27]Baylor College of Medicine Human Genome Sequencing Center, Houston, TX, USA. [28]Department of Epidemiology, Tulane University, New Orleans, LA, USA. [29]Tulane University Translational Science Institute, Tulane University, New Orleans, LA , USA. [30]Division of Genomics and Data Science, Department of Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA. [31]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA. [32]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. [33]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [34]Department of Pediatrics, The Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA. [35]Department of Medicine, Cardiovascular Division, Washington University School of Medicine, St. Louis, MO, USA. [36]National Yang-Ming University School of Medicine, Taipei, Taiwan. [37]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [38]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. [39]Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, USA. [40]Survey Research Center, Institute for Social Research, University of Michigan 426 Thompson St, Room Ann Arbor, MI 48104, USA. [41]Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [42]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA. [43]University of Maryland Marlene and Stewart Greenebaum Comprehensive Cancer Center, Baltimore, MD, USA. [44]University of Washington Department of Biostatistics, Seattle, WA 98195, USA. [45]Inova Translational Medicine Institute (ITMI), Inova Health Systems, Falls Church, VA, USA.
*Corresponding author. Email: ssunyaev@rics.bwh.harvard.edu
†These authors contributed equally to this work.
‡https://www.nhlbiwgs.org/topmed-banner-authorship; see "Additional Authors from the Trans-Omics for Precision Medicine Program" for full banner author list (excluding primary authors above).

whereas for others it is nontranscribed. As a consequence of TCR, some genic regions display depletion of A>G mutations, and others display depletion of the complementary T>C mutations (Fig. 1C).

For a strand-dependent mutational mechanism, our statistical procedure would infer two independent components with reverse-complementary spectra (Fig. 1C and fig. S1B). Following the example above, the intensity of A>G mutations in one component would be identical to the intensity of T>C in the other. By contrast, a strand-independent mechanism would correspond to a single self-complementary component (the intensity of A>G would be identical to the intensity of T>C). All biologically meaningful components would either be self-complementary or arise in mutually complementary pairs. We disregard spurious processes not conforming to this complementarity rule using the "reflection test" (Fig. 1, C and D) (*8*).

We applied our method to 292 million very rare (allele frequency below 10^{-4}) single-nucleotide variants (SNVs) from 42,813 individuals in the TOPMed freeze 5 (*9*). We applied a statistical correction (*8*) to account for multiple independent mutations that occurred in the same site (*10*). This correction increased the estimated rate of CpG>TpG mutations by 1.8-fold (fig. S2). To capture the regional variation, we binned the genome into non-overlapping windows of 10 kb (fig. S3C), the

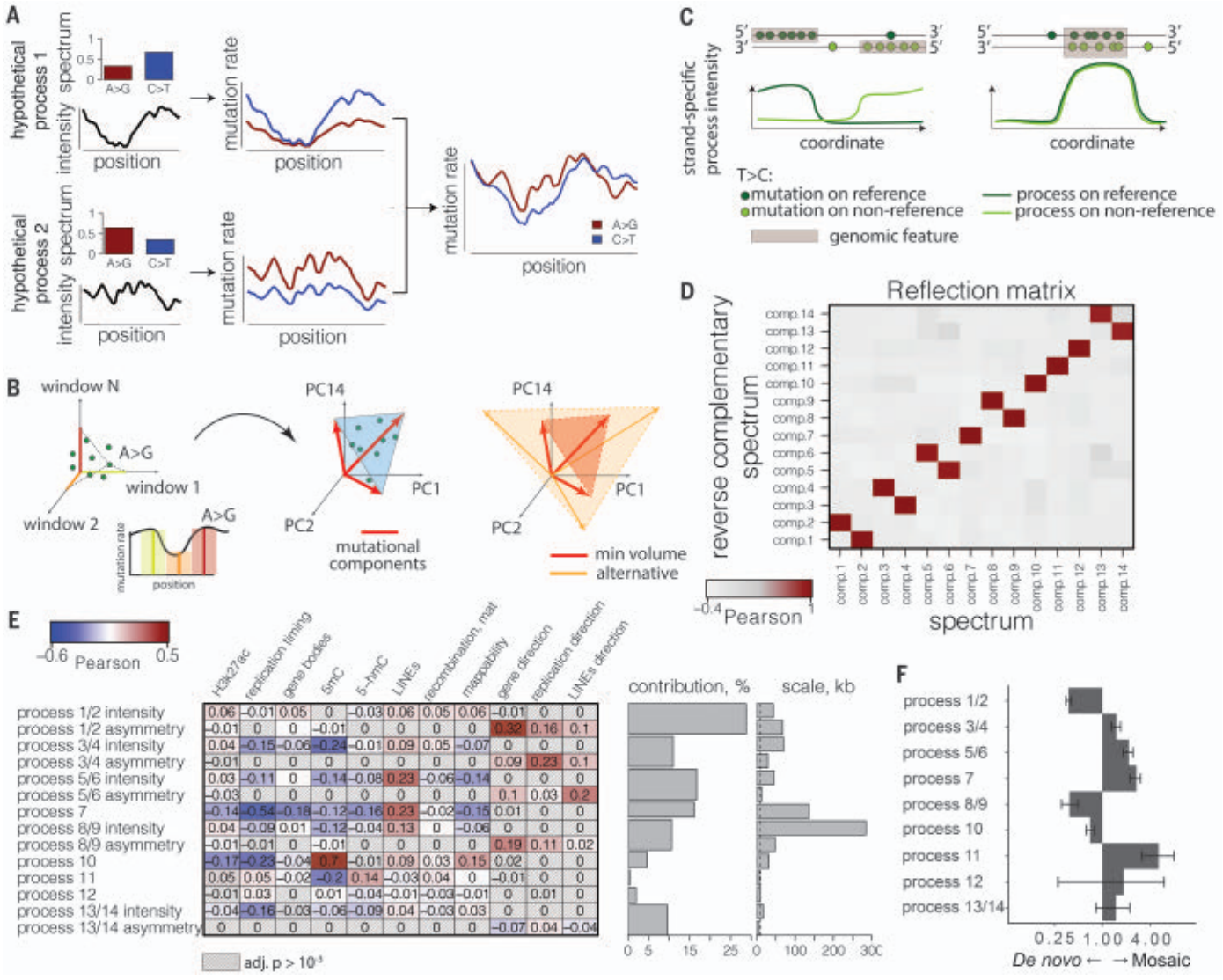


Fig. 1. Inference of spatially varying mutational processes in the human germ line. (**A**) Mutational data are modeled as the sum of processes defined by spectra and positional intensities. Two hypothetical processes (left) produce rates of the two mutation types (middle) that together generate the data (right). (**B**) Inference steps for volume-regularized NMF (vrnmf): Rates of mutation types in each locus (left) are represented in the low-dimensional space [principal components (PCs), middle]. In that space, vrnmf searches for the cone of minimum volume containing all the data; standard NMF identifies any cone containing the data (right). (**C**) A strand-independent process (left) has equal rates of mutations on the reference strand and complementary mutations on the nonreference strand at each locus. A strand-dependent mutational process (right) has unequal complementary mutation rates. (**D**) Reflection matrix reveals strand dependency of processes. Correlation of spectrum of one mutational component with reverse complementary spectrum of another component separates the components into self-correlated and mutually correlated pairs. (**E**) Left: Correlations of process intensities with genomic features. For strand-dependent processes, intensity is the sum of the two components, and asymmetry is the difference. Shaded correlations have Bonferroni-corrected p value > 0.001. High values of the replication timing track correspond to early replicating regions. Middle: Fraction of mutations contributed by the process. Right: Spatial scale of intensities. (**F**) Ratios between contributions of the processes to de novo versus early zygotic mutations (*8*).

scale that maximizes the number of reflected components (fig. S1H).

Application of vrnmf to this dataset identifies 14 components passing the "reflection test" and robust to resampling. They are reproduced in gnomAD (*10*) and recapitulated in de novo mutation data (figs. S1, G and L, and S4E) (*11*, *12*). The 14 components correspond to nine processes, five strand dependent (represented by two components) and four strand independent (Fig. 1D and fig. S1L). Eight processes correlate exclusively with one or two genomic features, including gene bodies, replication timing, direction of replication, and chromatin accessibility (Fig. 1E and fig. S4A). Processes have, on average, 40% higher correlation with genomic features than with any individual trinucleotide mutation type (fig. S4, B and C). This is especially notable given that the inference was solely based on mutation density.

Up to 5% of heritable mutations arise during early mosaic divisions (*13*). Intensities of several processes differ between de novo germline mutations and early mosaic mutations (Fig. 1F).

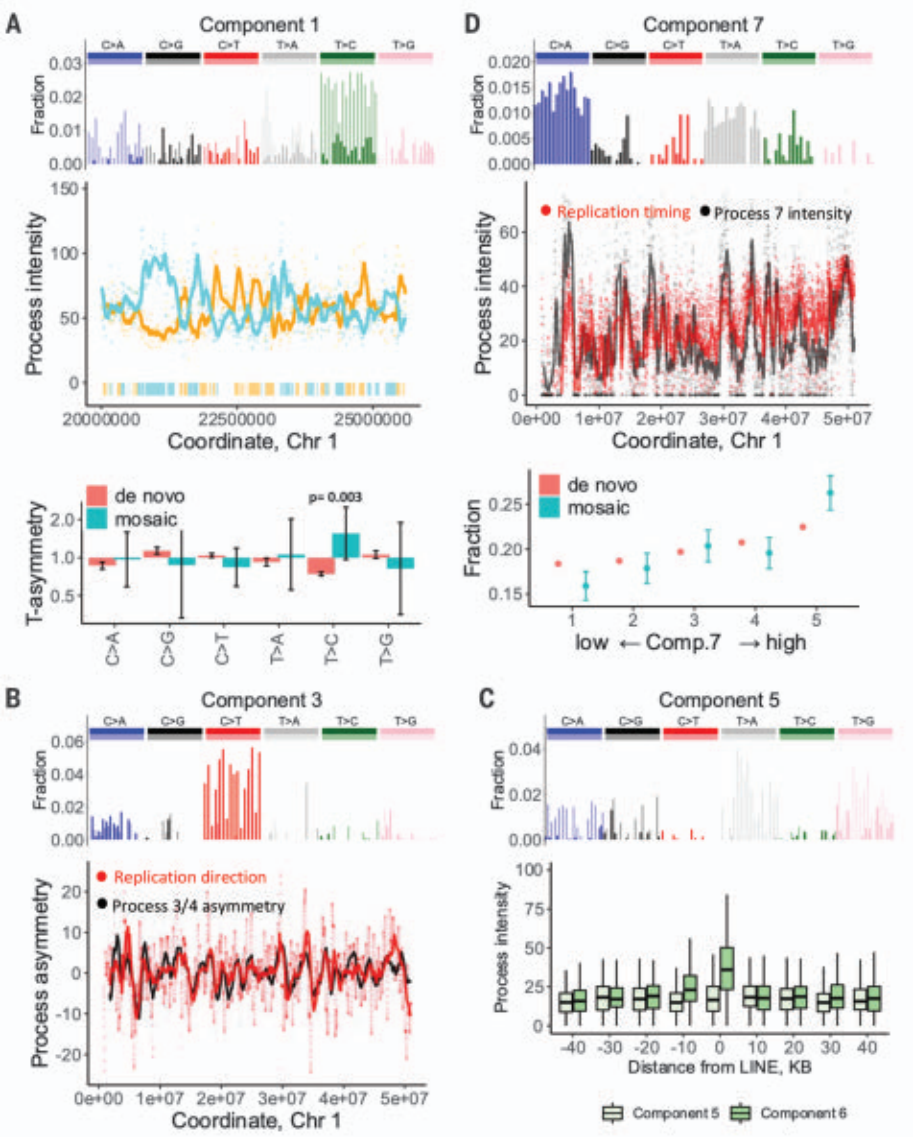


Fig. 2. Mutational processes are associated with distinct genomic features. (**A**) Top: The spectrum of component 1 (the reference strand is bright, and the nonreference strand is translucent). Mutation types are normalized to standard deviations Middle: Example of intensities of components 1 and 2. The bars depict gene bodies colored by direction of transcription. Bottom: Transcriptional asymmetry of de novo and mosaic mutations. (**B**) Top: The spectrum of component 3. Bottom: The association between asymmetry of process 3/4 and direction of replication. (**C**) Top: The spectrum of component 5. Bottom: Intensities of component 5 and component 6 near LINEs (template strand). (**D**) The spectrum of component 7 (top), and its association with replication timing (middle). (Bottom) Fractions of de novo and mosaic non-CpG mutations as a function of activity of process 7.

Broadly, mutations arise from replication errors or from DNA damage. Bulky damage is resolved in a strand-specific manner. TCR removes lesions on the transcribed strand (*2*), and replication converts damaged nucleotides to mutations more frequently on the lagging strand.

Strand-dependent process 1/2 is represented by mutually complementary components 1 and 2 (Figs. 1D and 2A). The strand asymmetry, measured as the difference between intensities of components 1 and 2, correlates with the directions of both transcription (r = 0.32) and replication (r = −0.15) and with gene expression (table S1). Process 1/2 correlates with the experimentally obtained TCR activity (*15*) (fig. S5A). We conclude that this process is a footprint of the asymmetric resolution of bulky DNA damage. Process 1/2 has reduced intensity in early development, and transcriptional asymmetry of mosaic A>G/T>C mutations has the opposite direction (Figs. 1F and 2A).

Strand-dependent process 3/4 likely captures asymmetric replication errors. Its asymmetry correlates with the direction of replication (Fig. 2B; r = 0.34 at the optimal 100-kb scale, table S2). This process is unlikely to be mediated by bulky DNA damage, because the correlation with direction of transcription is small. We hypothesize that process 3/4 reflects differential replication fidelity between leading and lagging strands (*1*), offering the first probable footprint of replicative errors.

Process 5/6 has an elevated intensity on nontranscribed strands of L1PA long interspersed nuclear element (LINE) repeats (Fig. 2C and fig. S5E). The 2% of the genome with the strongest asymmetry of process 5/6 are fourfold enriched with L1PA repeats but not with other LINEs (fig. S5D).

Strand-independent process 7 closely tracks replication timing (RT) (r = 0.64 at the optimal 500-kb scale) (Fig. 2D). The modest association of mutation rate with RT has been long known (*16*, *17*). It is stronger for transversions (*17*), especially C>A (*18*). The intensity of process 7 increases by 4.4-fold from the earliest to latest RT decile, whereas the rate of transversions increases by just 20% (fig. S5C). Process 7 is substantially more active in early development (Figs. 1F and 2D).

Strand-dependent process 8/9 is dominated by C>G transversions. It is characterized by local spikes totaling 264 Mb (Fig. 3, A to C); just 10% of the genome harbors 67% of clustered de novo mutations of maternal origin and includes all known (*19*) and many new regions of accelerated maternal mutagenesis (Fig. 3D and tables S3 and S4).

Process 8/9 displays a 50 to 200% increase in the rate of C>G mutations on the nontranscribed strand compared to gene flanks (Fig. 3, E to G, and figs. S6 and S7A). This effect is especially pronounced for long, fragile

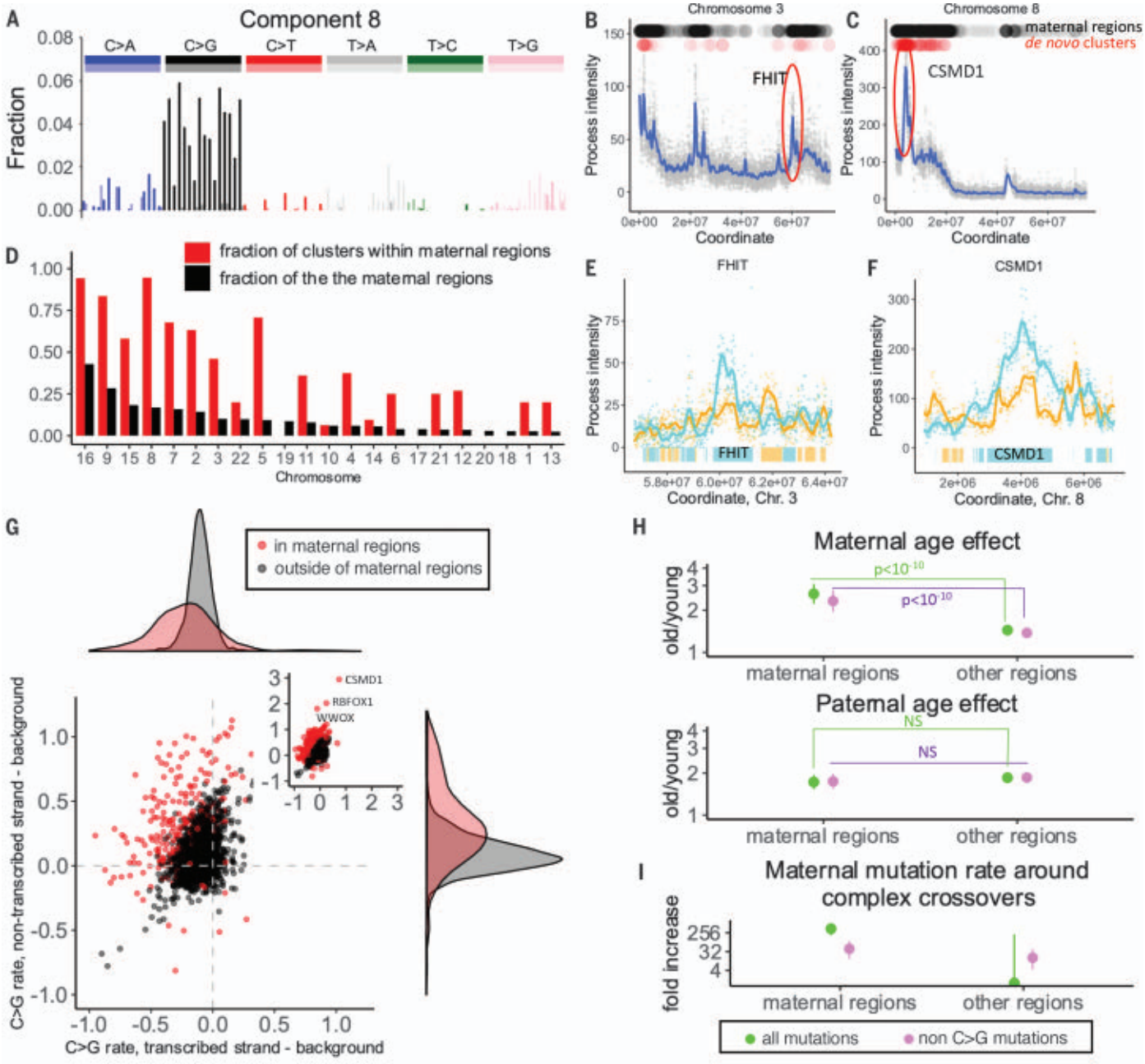


Fig. 3. Oocyte-specific mutational process. (**A**) The spectrum of component 8. (**B** and **C**) Examples of spikes of process 8/9 (black dots) alongside de novo maternal clustered mutations (*11*) (red dots). (**D**) Enrichment of maternal clustered de novo mutations (*11*) in spikes of process 8/9. (**E** and **F**) Spikes of process 8/9 around *FHIT* and *CSMD1* on nontranscribed strands. The bars depict gene bodies colored by direction of transcription. (**G**) C>G mutation rate on transcribed or nontranscribed strands compared to that on 100-kb flanks. (**H**) Ratio of parent-specific de novo mutation rates between the first and the last parental age quartiles. (**I**) Fold change in maternal de novo mutation rate in 100-kb windows around complex crossovers.

genes (*WWOX*, *RBFOX1*, *CSMD1*, *FHIT*, *SDK1*) covered by spikes of the process. We interpret this as transcription-associated mutagenesis in oocytes that is possibly induced by localized susceptibility to DNA damage (*20*). Mutations in spikes of process 8/9 show a much stronger effect of maternal age than the remaining genome (Fig. 3H).

Accumulation of maternal mutations with age in nondividing oocytes cannot be mediated by replication. Literature favors resolution of double-strand breaks (DSBs) as a likely mechanism (*11*, *19*, *21*). Complex crossovers in spikes of process 8/9 have 399-fold elevated C>G mutation rates (Fig. 3I), in line with (*11*). However, all complex crossovers contribute only 10 out of 507 additional C>G mutations in spikes of process 8/9 in the de novo mutation dataset (*11*), suggesting that this is an important but not a major mechanism.

Process 10, characterized by CpG transitions, is known to be mediated by methylcytosine deamination or by erroneous replication over methylcytosine (*8*, *22*) (Fig. 4, A and B, and fig. S8A).

Process 11 is represented by CpG transversions (Fig. 4C). It is likely a footprint of enzymatic demethylation, which proceeds through hydroxymethylated cytosines and abasic sites as intermediates (*23*). Unfinished

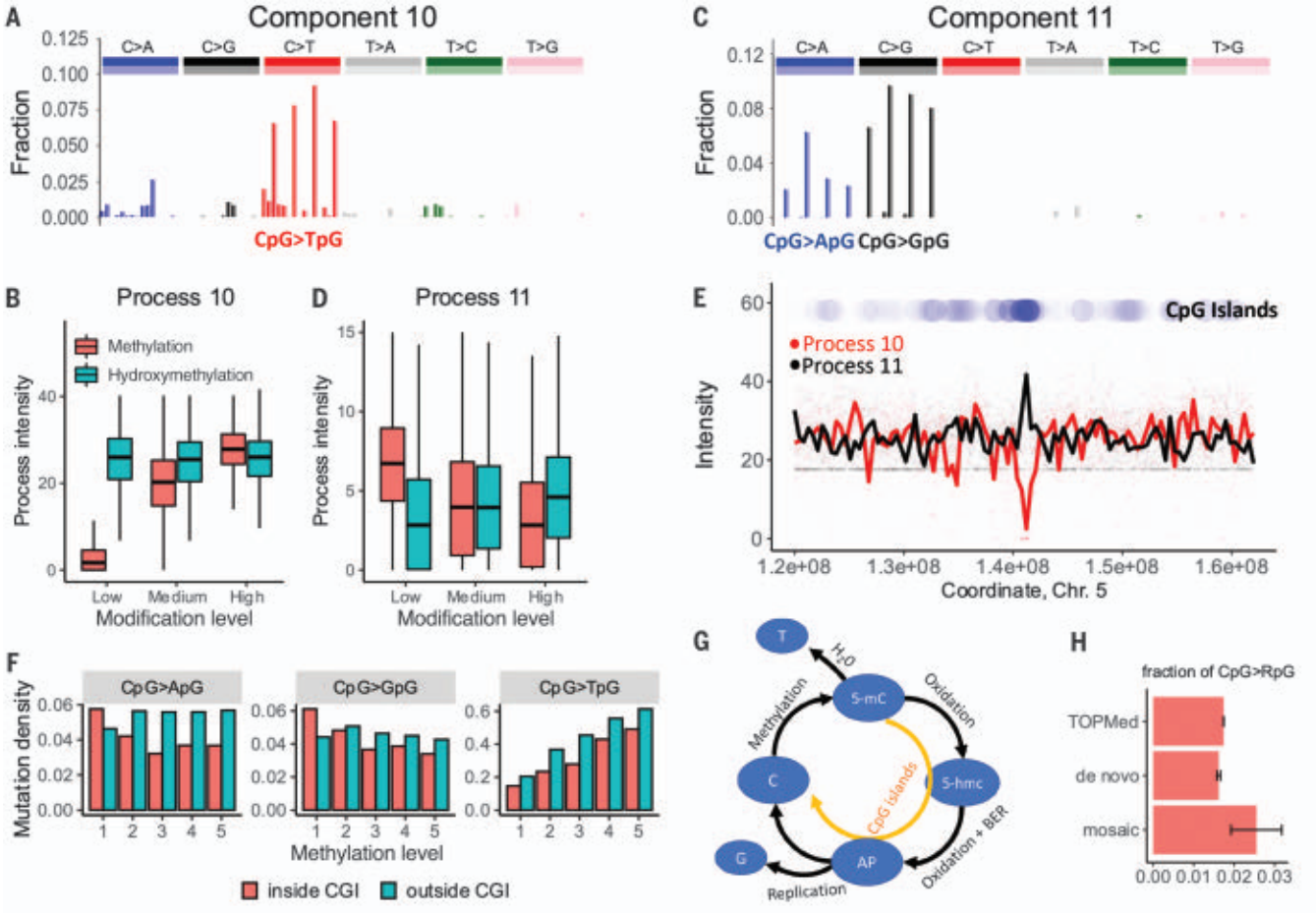


Fig. 4. Cytosine deamination and cytosine demethylation. (**A** and **C**) Spectra of components 10 and 11. (**B** and **D**) The intensity or processes 10 (C) and 11 (D) as a function of methylation and hydroxymethylation. (**E**) Process 10 increases and process 11 decreases in CpG islands (CGI). (**F**) Dependency of CpG mutations on methylation within and outside CGI. (**G**) Mechanisms suggested for processes 10 and 11. Oxidation of methylcytosine (5-mC) leads to hydroxymethylcytosine (5-hmC), which is removed by glycosylase, leaving an abasic site (AP). If not repaired prior to replication, AP sites cause CpG>GpG or CpG>ApG mutations. (**H**) Fraction of CpG transversions among mosaic mutations, de novo mutations, and rare polymorphisms.

repair of abasic sites results in CpG transversions (*24*). Process 11 negatively correlates with cytosine methylation (*25*) only in demethylated CpG islands and positively correlates with hydroxymethylation (*26*) (Fig. 4, D to G, and fig. S8). Process 11 has an increased activity in early mosaic mutations (Figs. 1F and 4H), likely driven by the demethylation wave in the early zygote (*26*).

The remaining unexplained processes 12 and 13/14 (figs. S9 and S10) are responsible for a small fraction of mutations (Fig. 1E).

Subsampling suggests no statistical signs of saturation for the number of detectable processes with respect to sample size (figs. S1, I and J, and S4D). Owing to limited power, the contributions of known mechanisms such as recombination are not observed in our analyses.

REFERENCES AND NOTES

1. T. A. Kunkel, D. A. Erie, *Annu. Rev. Genet.* **49**, 291–313 (2015).
2. J. A. Marteijn, H. Lans, W. Vermeulen, J. H. J. Hoeijmakers, *Nat. Rev. Mol. Cell Biol.* **15**, 465–481 (2014).
3. L. B. Alexandrov *et al.*, *Nature* **578**, 94–101 (2020).
4. K. Harris, J. K. Pritchard, *eLife* **6**, e24284 (2017).
5. H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, S. H. Jensen, *Comput. Intell. Neurosci.* **2008**, 764206 (2008).
6. X. Fu, K. Huang, N. D. Sidiropoulos, W.-K. Ma, *IEEE Signal Process. Mag.* **36**, 59–80 (2019).
7. A. M. S. Ang, N. Gillis, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **12**, 4843–4853 (2019).
8. See materials and methods.
9. D. Taliun *et al.*, *Nature* **590**, 290–299 (2021).
10. K. J. Karczewski *et al.*, *Nature* **581**, 434–443 (2020).
11. B. V. Halldorsson *et al.*, *Science* **363**, eaau1043 (2019).
12. J.-Y. An *et al.*, *Science* **362**, eaat6576 (2018).
13. T. A. Sasani *et al.*, *eLife* **8**, e46922 (2019).
14. V. B. Seplyarskiy *et al.*, *Nat. Genet.* **51**, 36–41 (2019).
15. S. Adar, J. Hu, J. D. Lieb, A. Sancar, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2124–E2133 (2016).
16. J. A. Stamatoyannopoulos *et al.*, *Nat. Genet.* **41**, 393–395 (2009).
17. A. Koren *et al.*, *Am. J. Hum. Genet.* **91**, 1033–1040 (2012).
18. I. Agarwal, M. Przeworski, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 17916–17924 (2019).
19. J. M. Goldmann *et al.*, *Nat. Genet.* **50**, 487–492 (2018).
20. S. Jinks-Robertson, A. S. Bhagwat, *Annu. Rev. Genet.* **48**, 341–359 (2014).
21. Z. Gao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9491–9500 (2019).
22. R. C. Poulos, J. Olivier, J. W. H. Wong, *Nucleic Acids Res.* **45**, 7786–7795 (2017).
23. X. Wu, Y. Zhang, *Nat. Rev. Genet.* **18**, 517–534 (2017).
24. K. Chan, M. A. Resnick, D. A. Gordenin, *DNA Repair (Amst.)* **12**, 878–889 (2013).
25. F. Supek, B. Lehner, P. Hajkova, T. Warnecke, *PLOS Genet.* **10**, e1004585 (2014).
26. H. Bagci, A. G. Fisher, *Cell Stem Cell* **13**, 265–269 (2013).
27. solrust, pkharchenko, *hms-dbmi/spacemut: Description of the data and code*, Zenodo (2021); https://doi.org/10.5281/zenodo.4494404.
28. solrust, pkharchenko, *kharchenkolab/vrnmf: Volume-regularize NMF*, Zenodo (2021); https://doi.org/10.5281/zenodo.4495386.

ACKNOWLEDGMENTS

We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. **Funding:** This work was supported by NIH grants R35GM127131, R01MH101244, U01HG009088, and R01 HG010372. R.A.S and P.V.K. were supported by NHLBI (R01HL131768). S.R. was supported by NIH R35HL135818. Whole-genome sequencing (WGS) for the Trans-Omics for Precision Medicine (TOPMed) program was supported by the National Heart, Lung, and Blood Institute (NHLBI). Specific funding sources for each study and genomic center and authors' contribution to dataset creation are given in supplementary text 1 and table S5. Centralized read mapping and genotype calling, along with variant quality metrics and filtering, were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the

TOPMed Data Coordinating Center (3R01HL-120393-02S1; contract HHSN268201800001I). **Author contributions:** V.S.B. and R.A.S. conceived the idea for the project and performed data analysis. S.S. and P.V.K. supervised the study, E.M.K. developed the correction for recurrent mutations, and R.J.M. carried out analysis of TET and DNMT3B binding. V.S.B., R.A.S., R.J.M., and S.S wrote the manuscript. The contributions of other authors are listed in supplementary text 1 and table S5. **Competing interests:** We declare no conflicts of interest. L.S.E. is currently an employee of Bristol Myers Squibb. Bristol Myers Squibb had no role in the funding, design, conduct, and interpretation of this study. **Data and materials availability:** There are no big datasets associated with the paper; all code and necessary data are available at (*27*) (https://doi.org/10.5281/zenodo.4494404): https://github.com/hms-dbmi/spacemut/. The vrnmf R package is available at (*28*) (https://doi.org/10.5281/zenodo.4495386): https://github.com/kharchenkolab/vrnmf.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/1030/suppl/DC1
Materials and Methods
Supplementary Text
Additional acknowledgments and phs numbers
Figs. S1 to S10
Tables S1 to S5
References (*29*–*46*)
MDAR Reproducibility Checklist

31 December 2019; accepted 14 July 2021
Published online 12 August 2021
10.1126/science.aba7408

PALEOCLIMATE

Insolation triggered abrupt weakening of Atlantic circulation at the end of interglacials

Q. Z. Yin[1]*, Z. P. Wu[1,2], A. Berger[1], H. Goosse[1], D. Hodell[3]

Abrupt cooling is observed at the end of interglacials in many paleoclimate records, but the mechanism responsible remains unclear. Using model simulations, we demonstrate that there exists a threshold in the level of astronomically induced insolation below which abrupt changes at the end of interglacials of the past 800,000 years occur. When decreasing insolation reaches the critical value, it triggers a strong, abrupt weakening of the Atlantic meridional overturning circulation and a cooler mean climate state accompanied by high-amplitude variations lasting for several thousand years. The mechanism involves sea ice feedbacks in the Nordic and Labrador Seas. The ubiquity of this threshold suggests its fundamental role in terminating the warm climate conditions at the end of interglacials.

Abrupt climate changes and millennial climate variability have been identified in numerous paleoclimate records, particularly during the last glacial period. Increasingly, studies are showing that abrupt and millennial- to centennial-scale climate changes can also be found in much older times (*1*–*6*), are a persistent feature for at least the past few million years, and occur not only during glacial times but also during interglacials (*4*, *7*–*11*). Many interglacial periods of the late Pleistocene were terminated by abrupt cooling events that mark the reappearance of strong millennial variability (*9*, *12*–*16*).

During the last interglacial Marine Isotope Stage 5e (MIS-5e), the first abrupt cooling characterizing the end of the interglacial can be observed in the Greenland ice core records (*17*), which is defined as Greenland stadial 26 and occurred at about 119 thousand years before present (ka BP) (±2.5 ka). A similar abrupt cooling is also found in the $\delta^{18}O$ record (*5*) of the Iberian Margin at 122.4 ka BP (±4 ka) (Fig. 1), which is consistent with the Greenland ice core records considering the age uncertainty. In fact, this $\delta^{18}O$ record shows that an abrupt cooling occurred at the end of each interglacial of the past 800 ka (Fig. 1), indicating that this is a general feature for all interglacials during the late Pleistocene. After the abrupt cooling, the climate enters into a cold state characterized by increased millennial-scale fluctuations.

The observation that these strong abrupt coolings occurred when climate was still in a warm interglacial state is puzzling and its cause remains uncertain. Astronomically induced long-term variation of insolation is one of the most important external forcings of the climate system. It acts as a pacemaker of climate change, driving the behavior of different components of the climate system. Because it varies slowly on the order of tens of thousands of years, it is often ignored in explaining abrupt climate changes on shorter time scales (10^2 to 10^3 years). To investigate the possible influence of the slow variation of insolation on interglacial climate, we performed transient climate simulations for all interglacials of the past 800 ka with an atmosphere-ocean-vegetation coupled model, LOVECLIM1.3 (see the supplementary materials). We considered 11 interglacial substages (*18*): MIS-1, MIS-5e, MIS-7ac, MIS-7e, MIS-9e, MIS-11c, MIS-13a, MIS-15a, MIS-15e, MIS-17c, and MIS-19c. To investigate the influence of insolation alone, only astronomically induced insolation was initially allowed to vary with time, and CO_2 concentration was fixed to a typical interglacial level of 280 parts per million volume (ppmv). In a second set of simulations, the impacts of both CO_2 and insolation changes were investigated.

During MIS-5e, in response to the variation of insolation alone, our simulated sea surface temperature (SST) in the northern North Atlantic varies slowly and follows closely the variation of summer insolation (Fig. 2C). However, an abrupt SST decrease (1.4°C in ~50 years) occurs at 119.4 ka BP and is followed by rapid and large oscillations over the next 7000 years with a much cooler mean climate state. Similar oscillations happen in the simulated surface air temperature (SAT), precipitation, and vegetation in different latitudes, especially in the Northern Hemisphere (NH) (fig. S1), indicating that these abrupt oscillations are a widespread phenomenon.

During the abrupt cooling at 119.4 ka BP, the oceans became cooler in the NH, especially in the northern North Atlantic, where the cooling could reach 1° to 5°C (Fig. 2D), and they became slightly warmer in the Southern Hemisphere (SH). The change in SAT (Fig. 2E) had a similar pattern but with a larger amplitude than SST. Over the Labrador Sea and the Barents Sea, the SAT decreased by up to 6° and 4°C, respectively. It decreased by 5°C over southern Greenland. The opposite change in temperature in the two hemispheres is associated with an abrupt weakening of Atlantic meridional overturning circulation (AMOC) (a weakening of 30% in 50 years) (Fig. 2B), followed by large-amplitude oscillations with a mean state of much weaker AMOC. The mechanism by which AMOC influences the climate in both hemispheres has been discussed in numerous studies (*19*). Stronger AMOC transports more heat to the NH high latitudes, leading to a generally warmer NH and cooler SH and vice versa. The abrupt weakening of AMOC at 119.4 ka BP was responsible for a reduction of 22% in the northward heat transport across 40°N in the North Atlantic. Therefore, the key to understanding the insolation-induced abrupt changes in our simulated temperature and other components of the climate system is how insolation induces such abrupt changes in AMOC.

The occurrence of abrupt weakening of AMOC followed by high-amplitude oscillations driven by slow insolation changes only is

[1]Georges Lemaître Center for Earth and Climate Research, Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. [2]Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China. [3]Godwin Laboratory for Palaeoclimate Research, Department of Earth Sciences, University of Cambridge, Cambridge, UK.
*Corresponding author. Email: qiuzhen.yin@uclouvain.be

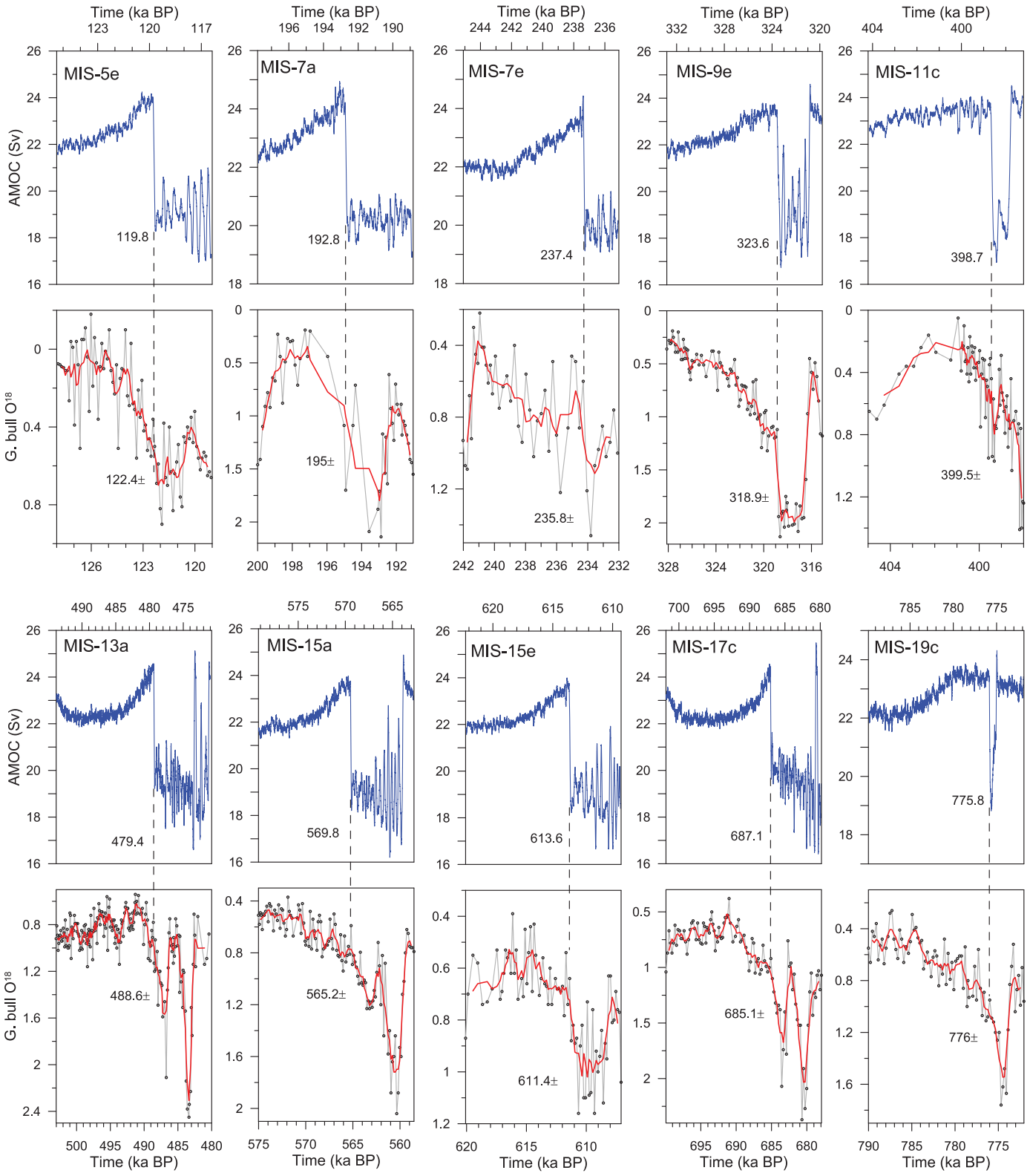


Fig. 1. The planktonic $\delta^{18}O$ record from the Iberian margin and the simulated AMOC intensity (blue). The $\delta^{18}O$ data are from ref. (*5*) augmented in this study by additional analyses over glacial inceptions. Gray dotted lines and red lines represent the $\delta^{18}O$ data and a five-point running mean of the data. The AMOC results are from the second set of simulations driven by both insolation and CO_2, and a 100-year running mean is applied. The dates of the abrupt cooling and abrupt weakening of AMOC are indicated. Given the age uncertainty (±4 ka) in the planktonic $\delta^{18}O$ record, the two datasets are plotted on their own time scales and are aligned at the time of the abrupt changes observed in each set of data. The date of the abrupt changes in the $\delta^{18}O$ data is defined as the time when the first sharp cooling occurs at the end of each interglacial period, corresponding to the age immediately before the first sharp increase of the $\delta^{18}O$ signal. G. bull., denotes the planktic foraminifera *Globigerina bulloides*.

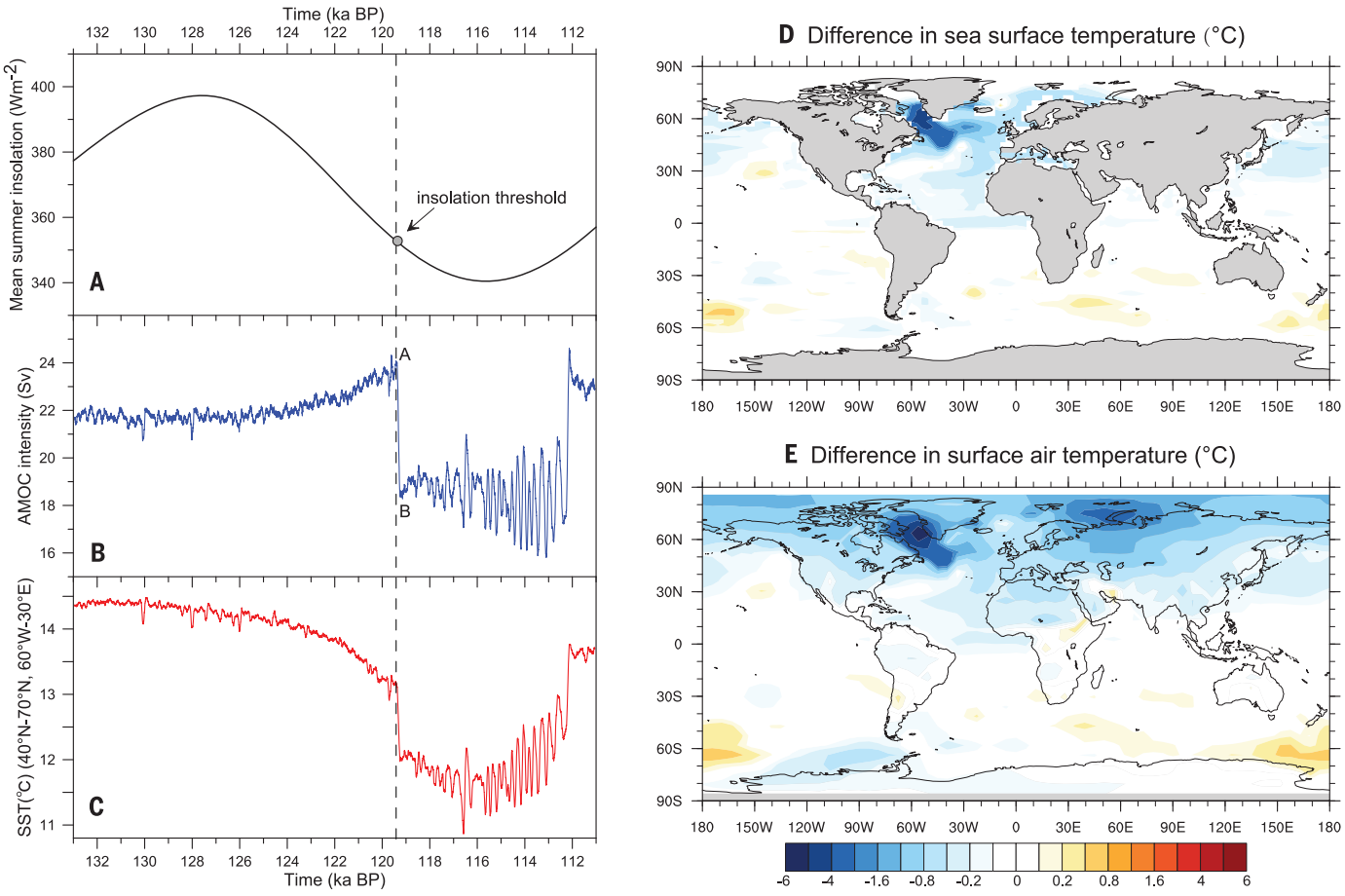


Fig. 2. Insolation-only induced variations in AMOC and temperature during MIS-5e. (**A**) Mean summer insolation averaged over the four latitudes 55°N, 65°N, 75°N, and 85°N (*23*); the mean insolation of the half-year NH astronomical summer is obtained by dividing the total irradiation received during the half-year summer (*24*) by the length of the half-year summer (*25*). (**B** and **C**) AMOC intensity (B) and annual mean SST (C) in the North Atlantic region. (**D**) Difference in annual mean SST between points "B" and "A." (**E**) Difference in annual surface air temperature between points "B" and "A." The results are from the simulation with only insolation varying and CO_2 fixed at 280 ppmv. A 100-year running mean is applied on the simulated AMOC and SST.

unexpected. It occurs not only during MIS-5e, but also during all the other interglacials except MIS-1, MIS-9e, and MIS-19 (fig. S2). The onset of the oscillations is always marked by a sharp weakening of AMOC that happens at the end of the interglacials. Moreover, the onset always occurs at times when NH summer insolation is decreasing and approaching a minimum. This leads to the proposition that a threshold of NH summer insolation may trigger an abrupt weakening of the AMOC. Our examination suggests that the threshold value is not exactly the same among the interglacials (fig. S3). This is not unexpected because the latitudinal and seasonal distribution of insolation is not exactly the same among the interglacials. Nevertheless, our results show that, to trigger the abrupt weakening of AMOC, the summer insolation must be sufficiently low, ranging in a narrow window between 352.1 Wm^{-2} for MIS-15e and 358.2 Wm^{-2} for MIS-7a (Fig. 3 and fig. S3).

The mean summer insolation is controlled by both obliquity and precession, with precession being the major controlling factor reflecting the important role of the length of the season. Our results (fig. S2) show that the favorable astronomical condition to trigger the abrupt weakening of AMOC is a combination of a high precession with June solstice occurring at aphelion (inducing long summers) and, at the same time, a relatively low obliquity (inducing low total summer irradiation). The insolation threshold values have a strong negative correlation with the precession index (fig. S3) because of the dominance of precession on the mean summer insolation. Obliquity also plays a role through its phase relationship with precession. For all the interglacials, when abrupt weakening and high-amplitude oscillations of AMOC occur, precession maxima and obliquity minima are in phase (fig. S2), leading to very low summer insolation in the NH. On the contrary, for some periods, e.g., around 210 and 495 ka BP, although precession reaches one of its maxima, abrupt oscillations do not occur because of the anti-phase relationship between precession and obliquity. In this case, high obliquity counteracts the effect of high precession, making the summer insolation insufficiently low. During MIS-1 and MIS-19, low eccentricity leads to small variations in precession and therefore there is no deep minima in summer insolation, explaining why insolation alone does not trigger abrupt oscillations during these two interglacials. Although MIS-11 also has low eccentricity similar to MIS-1 and MIS-19, the insolation is sufficiently low to trigger abrupt oscillations because of the low obliquity. In summary, to have abrupt weakening of AMOC, both low obliquity and high precession (occurrence of NH summer at aphelion and high eccentricity) are required.

Comparison between our two sets of simulations shows that lowered insolation is essential for triggering the abrupt changes of AMOC at the end of interglacials, and CO_2 only modulates slightly the effect of insolation (see the supplementary materials and fig. S2).

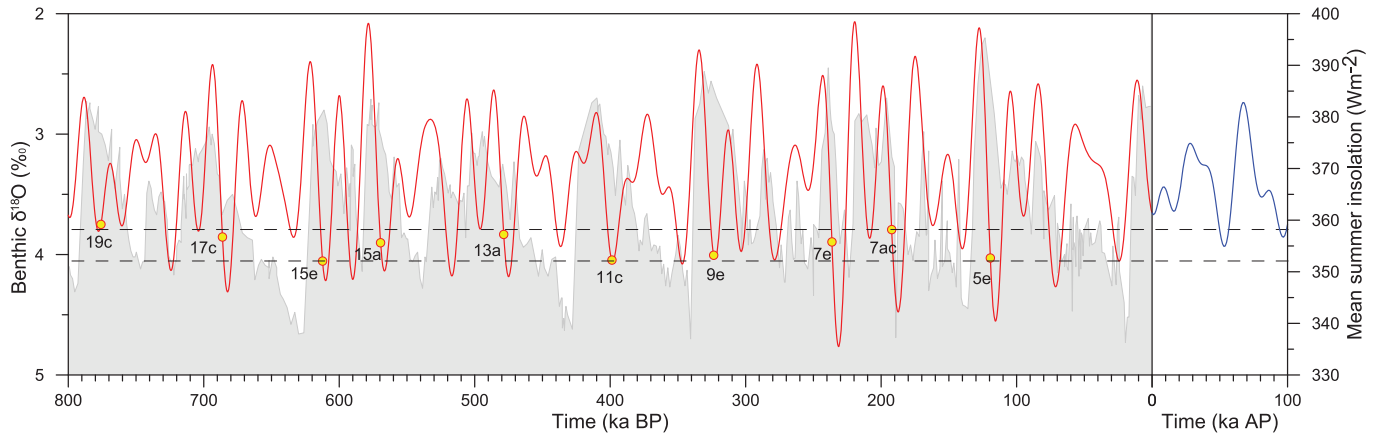


Fig. 3. Insolation threshold band through the past 800 ka and the future 100 ka. Red and blue curves are the mean summer insolation as defined in Fig. 2. Gray shaded curve is the benthic $\delta^{18}O$ at site U1385 (*5*) showing the glacial-interglacial cycles. The two horizontal dashed lines define the upper and lower insolation threshold values 358.2 and 352.1 Wm^{-2}, respectively, determined from the insolation-only simulations. Yellow dots indicate the times at the end of the interglacials when abrupt weakening of AMOC is triggered.

For MIS-9e and MIS-19, when insolation minima are insufficiently low to trigger the abrupt oscillations, the decrease of CO_2 reinforces the insolation-induced cooling and helps to trigger abrupt oscillations.

Figure 1 shows that under the combined influence of insolation and CO_2, our model simulates an abrupt weakening of AMOC at the end of all the interglacials of the past 800 ka. Their timings are in good agreement with those of the first abrupt cooling in the planktic $\delta^{18}O$ record from the Iberian Margin if age uncertainties are considered, except for MIS-13a, which is a controversial interglacial (*20*). After this time, the climate becomes unstable in both the simulation and the planktic $\delta^{18}O$ record. The first sudden cooling and reduction of AMOC at 119.8 ka BP in our last interglacial simulation also coincide with the first abrupt cooling observed at 119 ka BP (the stadial 26) in the Greenland ice core records (*17*) and ~118 ka BP in the SST in subpolar North Atlantic (*12*). It is worth noting that a series of cooling events occurring after MIS-5e have been observed in many records, and our simulated abrupt cooling corresponds to the very first cooling just after the interglacial optimum (see the supplementary materials). The SST reconstructions south of Greenland indicate a sudden cooling around 119, 238, 322 and 399 ka BP at the end of MIS-5e, MIS-7e, MIS-9e, and MIS-11c (*14*). At the end of MIS-11c, an abrupt weakening of the Asian monsoon is indicated around 397 ka BP (*9*), and an abrupt cooling around 396 ka BP is identified in the Lake Baikal region (*16*). All of these dates are consistent with the ages for each of the first abrupt AMOC weakening in our simulations, marking the end of the interglacials (Fig. 1) when age uncertainties are taken into account. The consistency between our model results and paleoclimate records suggests that the abrupt cooling at the end of interglacials occurred when insolation fell below the threshold value, and this consistency also supports the validity of the range of the insolation threshold values found in our study (Fig. 3).

In addition to the end of interglacials, there were other times in the past 800 ka when insolation fell below the threshold (Fig. 3). Most occurred during glacial times, and two during the stadials MIS-13b and MIS-15b. However, given the very different conditions during glacial times with much larger ice sheets and lower CO_2, the insolation threshold value determined from our interglacial climate simulations is most likely not the same for glacial times. Regarding the future, our simulation for the next 34 ka does not show abrupt change of AMOC (fig. S2). The insolation minima remain high during the next 50 ka because of a small eccentricity and reach the threshold only at 50 ka AP (Fig. 3). Such a long period without insolation reaching the threshold is exceptional compared with all the interglacials in the past 800 ka, including MIS-11 and MIS-19, which also have small eccentricity. It has been demonstrated that the current interglacial would be exceptionally long and the next glacial inception would occur only in 50,000 years (*21*).

To explain how slow variation of insolation alone triggers abrupt weakening and high-amplitude oscillations of AMOC, we use the example case of MIS-5e. The processes are similar for other interglacials. Figure 4A shows that the strength of the AMOC is anticorrelated with summer insolation on time scales of orbital variations. A previous study (*22*) showed that over the high-latitude oceans, summer insolation plays a key role not only in controlling summer SST and sea ice, but also during the winter through the remnant effects of summer insolation. Our analyses show that this antiphase relationship between summer insolation and the intensity of AMOC can be explained by the fact that when insolation decreases, the SST in the northern North Atlantic decreases and more sea ice is formed, leading to saltier surface water as a consequence of brine rejection. In addition, the net precipitation over the ocean south of Greenland decreases substantially, leading to additional salinity increase. Lower SST and saltier water combine to contribute to denser surface water, stronger convection, and stronger AMOC.

The smooth relationship between summer insolation and AMOC described above is disrupted by abrupt oscillations in AMOC as soon as insolation decreases below the threshold value. During the normal AMOC state (Fig. 4A, point A), deep mixing in the North Atlantic occurs at two centers, the southern Labrador Sea and the Nordic Sea (Fig. 4D). The intensity of AMOC is controlled by convection over these two regions. In the northern Nordic Sea, before 122 ka BP (Fig. 4B, point B), the convection intensifies smoothly with decreasing insolation. However, starting from 122 ka BP, it weakens smoothly despite continuously decreasing insolation and then, starting ~119 ka BP, multicentennial-scale, high-amplitude variations occur. In the Labrador Sea, the intensification of convection is accelerated from 122 ka BP but is suddenly shutdown at 119.4 ka BP (Fig. 4B, point C) and remains inactive over the next 7000 years. These results show that the abrupt weakening in AMOC at 119.4 ka BP is caused by the sudden shutdown of convection in the Labrador Sea, whereas the high-amplitude variability of AMOC between 119 and 112 ka BP is related to changes in the northern Nordic Sea.

The changes in the convection in the Labrador Sea and northern Nordic Seas are closely linked

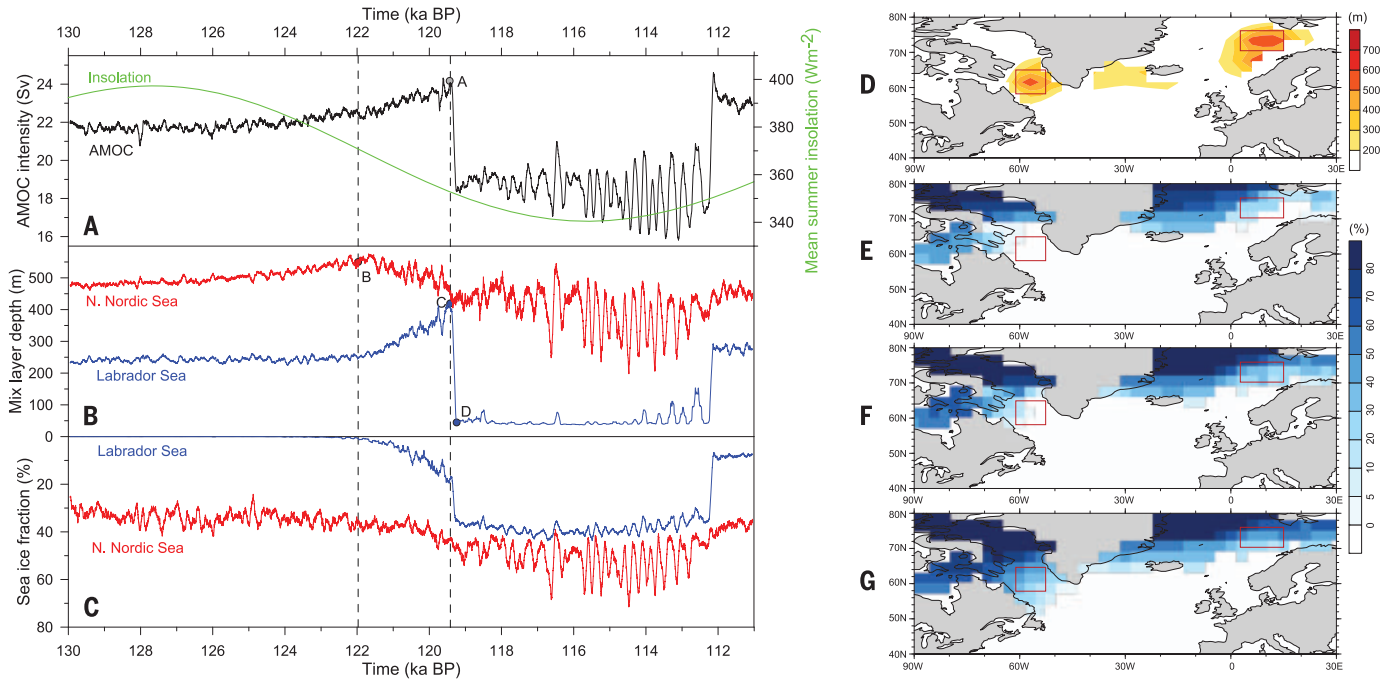


Fig. 4. Insolation-only induced variations in AMOC, the convection and sea ice in northern Nordic Sea and Labrador Sea during MIS-5e. (**A**) Mean summer insolation averaged over 55°N, 65°N, 75°N, and 85°N (green) and AMOC intensity (black). (**B**) Mix layer depth in the northern Nordic Sea (70°N to 78°N, 2.5°E to 17.5°E) (red) and the Labrador Sea (56°N to 66°N, 62°W to 52°W) (blue). (**C**) Sea ice fraction in the northern Nordic Sea (red) and the Labrador Sea (blue). (**D**) Mix layer depth (m) at time "A." (**E**), (**F**), and (**G**), Sea ice fraction at time "B," "C," and "D," respectively. A 100-year running mean is applied on the simulated results. The two rectangles in (D) to (G) indicate the convection center in the Labrador Sea and the northern Nordic Sea.

to the changes in sea ice over these regions. Over the southern Labrador Sea, there is no sea ice before 122 ka BP because of the relatively high summer insolation (Fig. 4, C and E). Because of the cooling induced by decreasing insolation, sea ice starts to develop there at 122 ka BP (Fig. 4C). The appearance of sea ice amplifies and accelerates the cooling over the Labrador Sea, which explains the accelerated deepening of convection after 122 ka BP (Fig. 4B up to point C). Sea ice continues to grow and starts to cover the center of convection in the Labrador Sea at 119.4 ka BP (Fig. 4F, point C), leading finally to a sharp shutdown of convection and a complete coverage of sea ice in the Labrador Sea in only a few decades (Fig. 4G).

In the northern Nordic Sea, before 122 ka BP, the convection is mainly controlled by the gradual cooling related to decreasing insolation, which explains the gradual intensification of convection. With insolation continuing to drop, cooling intensified, Arctic sea ice extended southwards and started to cover the convection center in the northern Nordic Sea by 122 ka BP (Fig. 4E), which explains the weakening of northern Nordic convection starting from this time (Fig. 4B). This lasts until 119 ka BP with no abrupt weakening, as seen in the Labrador Sea, but was immediately followed by high-amplitude, multicentennial oscillations.

These multicentennial oscillations in the convection of the northern Nordic Sea can be explained by the feedbacks between sea ice and ocean temperature. To illustrate this effect, we focus on three cycles. When sea ice increases over the northern Nordic Sea, the convection is weakened and becomes inactive when sea ice reaches its maximum and covers almost completely the convection center in northern Nordic Sea (fig. S4a, phase A, fig. S4b and fig. S4c. This convection-inactive period lasts for ~70 years. In the meantime, convection nevertheless remains active in the southern Nordic Sea (fig. S4c) because this region is not covered by sea ice. This explains why AMOC is not completely shut down even though the convection in both the Labrador Sea and the northern Nordic Sea is inactive during phase A. When sea ice covers completely the northern Nordic Sea, the subsurface ocean temperature increases (fig. S4a, phase A) because increased sea ice cover leads to reduced ocean heat loss to the atmosphere and consequently more heat stored in the ocean, resulting in subsurface warming. The ocean heat accumulates with time and melts sea ice and warms the air. The rapid release of heat from ocean to atmosphere after a rather long period of heat accumulation explains the sudden decline of sea ice at the end of phase A (fig. S4a). In the meantime, subsurface ocean temperature decreases rapidly. Once sea ice retreats from the convection site in the northern Nordic Sea, deep mixing starts again and convection is switched on (fig. S4a, phase B, and fig. S4d), which leads to the intensification of AMOC. The intensification of AMOC in turn enhances the northward oceanic heat transport, which explains the rapid subsurface warming in the northern Nordic Sea after the rapid cooling at the end of phase A. However, because insolation remains very low, the background climate is still too cold, and thus the Arctic sea ice extends again to the south and covers the convection center in the northern Nordic Sea. This leads to a new cycle of the multicentennial oscillations (fig. S4a).

The high-amplitude oscillations of AMOC only stop when insolation becomes high enough such that sea ice no longer covers the convection centers of the northern Nordic Sea and Labrador Sea. The duration of the high-amplitude oscillations differs among interglacials and ranges from 3100 years for MIS-11c to 8000 years for MIS-7ac and MIS-7e (fig. S2). The duration is related to the length of the intervals over which insolation is lower than the threshold value. During these periods, AMOC varies at multicentury scale with a major periodicity between 300 and 600 years.

The good agreement in the timing of the first cooling at the end of the interglacials between various paleoclimate records and our

model results supports an insolation threshold associated with the resumption of millennial variability near the end of all interglacial periods. The insolation-induced abrupt weakening of AMOC happens at the end of the interglacials, during which time the CO_2 concentration is still relatively high and ice sheets are relatively small. The millennial event triggered by declining insolation may very well lead to the end of interglacial conditions because it results in a strong abrupt cooling and a cold mean climate state lasting for several thousands of years in the northern middle-high latitudes and high snow accumulation over ice sheet nucleation regions (see the supplementary materials). These conditions are favorable for ice sheet growth that, in turn, amplifies the cooling and definitively terminates the warm interglacial period.

REFERENCES AND NOTES

1. J. Jouzel *et al.*, *Science* **317**, 793–796 (2007).
2. K. A. Gibson, L. C. Peterson, *Geophys. Res. Lett.* **41**, 969–975 (2014).
3. H. Cheng *et al.*, *Geophys. Res. Lett.* **39**, L01705 (2012).
4. B. Martrat *et al.*, *Science* **306**, 1762–1765 (2004).
5. D. A. Hodell *et al.*, *Global Planet. Change* **133**, 49–64 (2015).
6. B. Birner, D. A. Hodell, P. C. Tzedakis, L. C. Skinner, *Paleoceanography* **31**, 203–217 (2016).
7. E. Palumbo *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **383**, 27–41 (2013).
8. P. C. Tzedakis *et al.*, *Nat. Commun.* **9**, 4235 (2018).
9. X. N. Zhao *et al.*, *Geophys. Res. Lett.* **46**, 9949–9957 (2019).
10. E. V. Galaasen *et al.*, *Science* **367**, 1485–1489 (2020).
11. C. Nehrbass-Ahles *et al.*, *Science* **369**, 1000–1005 (2020).
12. D. W. Oppo, J. F. McManus, J. L. Cullen, *Quat. Sci. Rev.* **25**, 3268–3277 (2006).
13. Z. Mokeddem, J. F. McManus, D. W. Oppo, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11263–11268 (2014).
14. N. Irvalı *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 190–195 (2020).
15. E. V. Galaasen *et al.*, *Science* **343**, 1129–1132 (2014).
16. A. A. Prokopenko *et al.*, *Clim. Past* **6**, 31–48 (2010).
17. North Greenland Ice Core Project Members, *Nature* **431**, 147–151 (2004).
18. Past Interglacials Working Group of PAGES, *Rev. Geophys.* **54**, 162–219 (2016).
19. S. Rahmstorf, *Nature* **419**, 207–214 (2002).
20. Z. T. Guo, A. Berger, Q. Z. Yin, L. Qin, *Clim. Past* **5**, 21–31 (2009).
21. A. Berger, M. F. Loutre, *Science* **297**, 1287–1288 (2002).
22. Q. Z. Yin, A. Berger, *Clim. Dyn.* **38**, 709–724 (2012).
23. A. Berger, M. F. Loutre, *Quat. Sci. Rev.* **10**, 297–317 (1991).
24. A. Berger, M. F. Loutre, Q. Z. Yin, *Quat. Sci. Rev.* **29**, 1968–1982 (2010).
25. A. Berger, Q. Z. Yin, "Astronomical theory and orbital forcing," in *The Sage Handbook of Environmental Change*, J. A. Matthews *et al.*, Eds. (Sage, 2012), pp. 405–425.

ACKNOWLEDGMENTS

Computational resources have been provided by the supercomputing facilities of the Université Catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under convention 2.5020.11. Z.P.W. acknowledges the support of F.R.S.-FNRS under grant MIS F.4529.18 and the National Natural Science Foundation of China (grant nos. 41690114 and 41888101) for his doctoral study. Q.Z.Y. is a research associate and H.G. is a research director at F.R.S.-FNRS. **Funding:** This work is supported by the F.R.S.-FNRS under grant MIS F.4529.18. **Author contributions:** Q.Z.Y. designed the study and wrote the first draft of the paper. Q.Z.Y. and Z.P.W. performed model simulations and analyzed model results. D.H. provided the planktonic $\delta^{18}O$ data. All authors contributed to discussions and writing of the final paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The code for LOVECLIM1.3 is available at www.climate.be/loveclim. The planktonic $\delta^{18}O$ data used in this study are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/1035/suppl/DC1

Supplementary Text
Figs. S1 to S6
References (*26*–*38*)
Data S1

15 December 2020; accepted 29 July 2021
10.1126/science.abg1737

MICROBIOME

Enterococcus peptidoglycan remodeling promotes checkpoint inhibitor cancer immunotherapy

Matthew E. Griffin[1,2], Juliel Espinosa[1], Jessica L. Becker[1], Ji-Dung Luo[3], Thomas S. Carroll[3], Jyoti K. Jha[4], Gary R. Fanger[4], Howard C. Hang[1,2]*

The antitumor efficacy of cancer immunotherapy can correlate with the presence of certain bacterial species within the gut microbiome. However, many of the molecular mechanisms that influence host response to immunotherapy remain elusive. In this study, we show that members of the bacterial genus *Enterococcus* improve checkpoint inhibitor immunotherapy in mouse tumor models. Active enterococci express and secrete orthologs of the NlpC/p60 peptidoglycan hydrolase SagA that generate immune-active muropeptides. Expression of SagA in nonprotective *E. faecalis* was sufficient to promote immunotherapy response, and its activity required the peptidoglycan sensor NOD2. Notably, SagA-engineered probiotics or synthetic muropeptides also augmented anti–PD-L1 antitumor efficacy. Taken together, our data suggest that microbiota species with specialized peptidoglycan remodeling activity and muropeptide-based therapeutics may enhance cancer immunotherapy and could be leveraged as next-generation adjuvants.

Cancer immunotherapy harnesses the patient's immune system to impede tumor growth and has demonstrated clinical success against a variety of solid and hematological tumors (*1*–*3*). In particular, antibodies that target immune checkpoint inhibitor proteins such as CTLA-4, PD-1, and PD-L1 have been approved to treat a number of human cancers (*3*). However, patient response to immune checkpoint inhibitors is variable (*3*). The success of checkpoint blockade relies on numerous factors including mutational burden of the malignancy (*4*), successful tumor antigen presentation (*5*), recruitment and infiltration of lymphocytes (*6*), and signaling cues within the tumor microenvironment (*7*). Recently, the gut microbiota has emerged as a potent factor associated with the efficacy of anti–CTLA4, anti–PD-1, and anti–PD-L1 cancer immunotherapies (*8*–*14*). In animal and human cohorts, the presence of specific microbial species has correlated with the responsiveness of tumors to checkpoint blockade agents. The antitumor activity of these microbes was recapitulated in preclinical mouse models upon cohousing, fecal transplants, or direct inoculation, suggesting that the correlated microorganisms are direct causative agents of improved therapeutic response. Nevertheless, little is known about the molecular mechanisms by which immune modulatory microbes may behave.

Analyses of the commensal microbiome from human cohorts treated with immunotherapies targeting PD-1 have revealed that the bacterial genus *Enterococcus* is enriched in responsive patients (*12*, *13*). Although antibiotic-resistant strains of *E. faecium* and *E. faecalis* can be pathogenic (*15*), commensal strains of these bacteria have been used as probiotics in animals and humans (*16*). Recent studies have also suggested that *Enterococcus* species can trigger immune signaling pathways and modulate infection (*17*–*19*), autoimmunity (*20*), and graft-versus-host disease (*21*). These observations prompted our inquiry into whether specific *Enterococcus* species and strains can improve responses to checkpoint inhibitor immunotherapy.

To evaluate specific enterococci and their mechanism of action, we used mouse tumor models and oral administration of bacteria. Specific pathogen–free (SPF) C57BL/6 mice from The Jackson Laboratory were pretreated for 2 weeks with a broad-spectrum antibiotic cocktail (containing ampicillin, colistin, and

[1]Laboratory of Chemical Biology and Microbial Pathogenesis, The Rockefeller University, 1230 York Ave., New York, NY 10065, USA. [2]Departments of Immunology and Microbiology and Department of Chemistry, Scripps Research, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. [3]Bioinformatics Resource Center, The Rockefeller University, 1230 York Ave., New York, NY 10065, USA. [4]Rise Therapeutics, 1405 Research Blvd. Suite 220, Rockville, MD 20850, USA.
***Corresponding author. Email: hhang@scripps.edu**

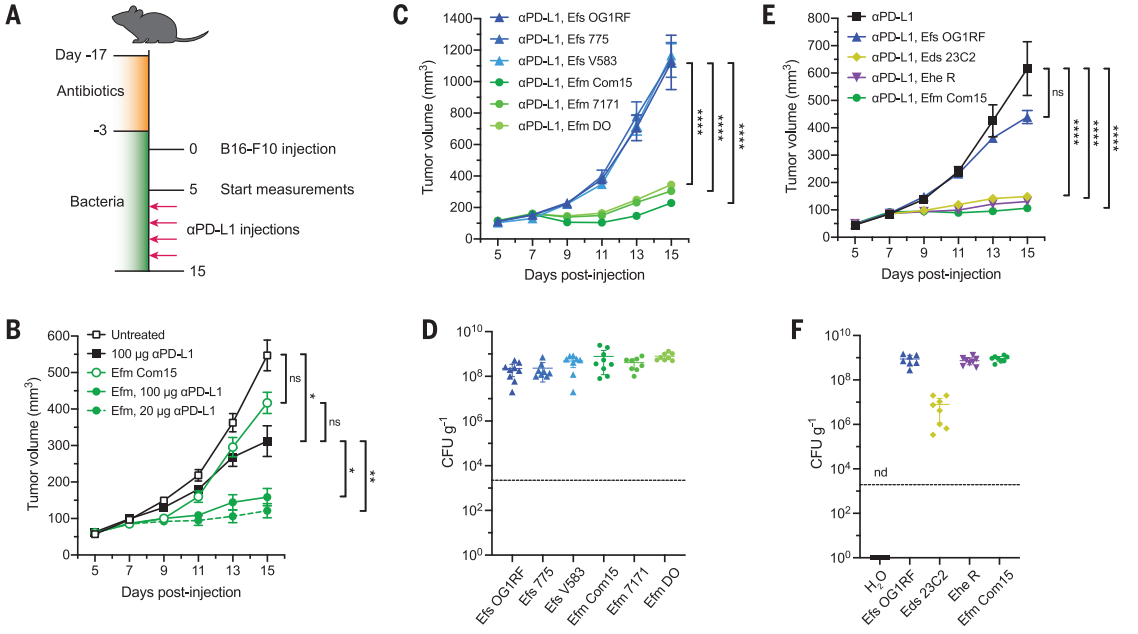


Fig. 1. Specific enterococci improve the efficacy of anti–PD-L1 cancer immunotherapy in the B16-F10 melanoma model. (**A**) Schematic of tumor growth model in SPF mice with antibiotic treatment and oral enterococci supplementation. Days are indexed based on day of tumor injection. Mice were provided water containing antibiotics ad libitum for 2 weeks, followed by water supplemented with the indicated enterococci for the remainder of the experiment. Animals were then subcutaneously implanted with B16-F10 melanoma cells, and tumor volume measurements started when the tumors reached ~50 to 100 mm^3 (day 5). Mice were treated with anti–PD-L1 by intraperitoneal injection every other day beginning 2 days after the start of tumor measurement. For all data except those shown in (B), 20 μg of anti–PD-L1 were used for each injection. (**B**) B16-F10 tumor growth in antibiotic-treated animals that were supplemented with or without *E. faecium* Com15 and treated with or without anti–PD-L1 starting on day 7, at doses indicated in the legend. $n = 7$ to 8 mice per group. (**C**) B16-F10 tumor growth in antibiotic-treated mice that were supplemented with the indicated *E. faecalis* and *E. faecium* strains and treated with anti–PD-L1 starting on day 7. $n = 7$ to 8 mice per group. (**D**) CFU analysis of *E. faecalis* and *E. faecium* strains in fecal samples harvested from animals treated as in (C). (**E**) B16-F10 tumor growth in antibiotic-treated mice that were supplemented with the indicated enterococci strains and treated with anti–PD-L1 starting on day 7. $n = 8$ to 9 mice per group. (**F**) CFU analysis of enterococci in fecal samples harvested from animals treated as in (E). nd, not detected. For (B), (C), and (E), data represent mean ± SEM. and were analyzed using a mixed effects model with Tukey's multiple comparisons post hoc test. *$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$; ns, not significant. For (D) and (F), each symbol represents one mouse. Dotted lines indicate the limit of detection (2000 CFU g^{-1}). Data represent means ± 95% confidence interval.

streptomycin) to clear resident microbial species that may confound effects caused by *Enterococcus* species and promote *Enterococcus* colonization. Animals were subsequently provided water supplemented with enterococci. Supplemented animals were then subcutaneously implanted with B16-F10 melanoma cells and treated with anti–PD-L1 (Fig. 1A). We first focused on the two most common *Enterococcus* species in the human gut microbiota: *E. faecium* (Efm) and *E. faecalis* (Efs). We found that supplementation with the human commensal *E. faecium* strain Com15 without therapeutic intervention did not alter tumor growth (Fig. 1B). However, treatment of supplemented animals with anti–PD-L1 immunotherapy at both high and low doses enabled a significant decrease in tumor size compared with anti–PD-L1 treatment alone. We selected the low anti–PD-L1 dose protocol going forward and compared the activity of *E. faecium* with *E. faecalis* across multiple strains of each species—including human-isolated, reference type, and multidrug-resistant strains—to ascertain whether the observed synergistic activity was specific to *E. faecium* Com15. Inhibition of tumor growth was observed for all three tested isolates of *E. faecium* but not for any analyzed strain of *E. faecalis* (Fig. 1C). To ensure that these effects were not due to differences in bacterial load, fecal samples were plated onto *Enterococcus*-selective medium and enumerated. All bacterial strains yielded similar numbers of colony forming units (CFUs) (Fig. 1D), indicating that the observed antitumor activity of the *E. faecium* strains was not due to differences in the amount of bacteria in the gut. In addition to *E. faecium*, 16*S* ribosomal RNA (rRNA) analysis of gut microbiota from responsive human patients also uncovered the enrichment of other *Enterococcus* species (*12*). We found that antitumor activity was conserved across multiple species including *E. durans* (Eds), *E. hirae* (Ehe), and *E. mundtii* (Emi) (Fig. 1E and fig. S1A). CFU analysis of colonized animals again revealed that antitumor activity did not correlate with bacterial load (Fig. 1F and fig. S1B).

To explain the species-specific differences we observed across the *Enterococcus* genus, we examined possible sources of their immunomodulatory activity. Our previous work indicated that *E. faecium* has distinctive peptidoglycan composition and remodeling capabilities to enhance host tolerance to enteric pathogens (*17*–*19*). To compare peptidoglycan composition across *Enterococcus* species, we isolated sacculi from each species and analyzed the digested peptidoglycan fragments by high-performance liquid chromatography–mass spectrometry (HPLC-MS). All four of the immunotherapy-active enterococci (*E. faecium*, *E. durans*, *E. hirae*, and *E. mundtii*) showed similar peptidoglycan fragment patterns that were distinct from those of the nonactive species, *E. faecalis* and *E. gallinarum* (Egm) (fig. S2A). Additionally, all four immunotherapy-active species showed an abundance of smaller, non–cross-linked peptidoglycan fragments, suggesting higher levels of peptidoglycan remodeling and turnover. The modification of peptidoglycan stem peptides is catalyzed by conserved families of amidases and peptidases (*22*), so we examined *Enterococcus* genomic assemblies for characteristic expression of peptidoglycan remodeling enzymes (fig. S3 and tables S1 and S2). Through this analysis, we identified a cluster of NlpC/p60 hydrolases that were highly conserved through all active enterococci (Fig. 2A). This group of related enzymes contained the peptidoglycan hydrolase

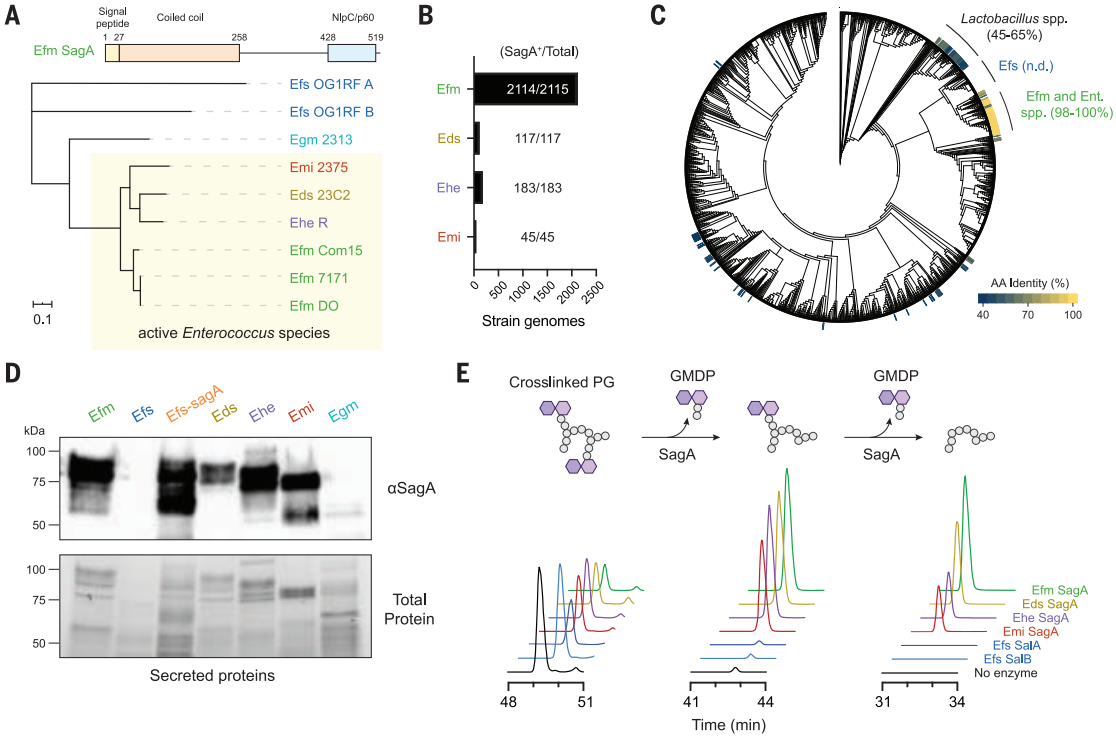


Fig. 2. Protective enterococci express and secrete active orthologs of the peptidoglycan NlpC/p60 hydrolase SagA. (**A**) Domain structure and unrooted phylogenetic clustering of putative SagA ortholog protein sequences identified by global peptidoglycan peptidase analysis of *Enterococcus* species and strains along with the closest entries from *E. gallinarum* and *E. faecalis* based on IQ-Tree analysis. Numbers above each domain denote amino acid residue boundaries. Active strains are indicated by the yellow box. Scale bar represents sequence distance. (**B**) Bar plot and quantification of *Enterococcus* genomes containing SagA orthologs. (**C**) Cladogram of Human Microbiome Project isolates organized by 16S rRNA homology with heatmap indicating amino acid (AA) sequence identity of putative SagA orthologs. nd, not detected; Ent. spp., *Enterococcus* strains without an assigned species name. (**D**) Western blot detection of secreted SagA orthologs harvested from overnight cultures of the indicated enterococci using antiserum raised against *E. faecium* Com15 SagA. Bottom panel shows total protein loading. Numbers indicate estimated molecular weight in kilodaltons (kDa). (**E**) In vitro activity of purified SagA orthologs. Data are shown as extracted LC-MS ion chromatograms of a cross-linked peptidoglycan substrate and two iterative hydrolysis products after incubating a mixture of peptidoglycan fragments with purified SagA orthologs from the indicated species for 16 hours at 37°C. Peak heights are shown as relative intensity of ion abundance, and all chromatograms are shown on the same scale.

secreted antigen A (SagA) from *E. faecium*, which improves host immunity against enteric infections (*17*–*19*). Using primary sequence alignment, we found that the putative SagA orthologs from other active enterococci were highly homologous, with >90% sequence identity within the C-terminal NlpC/p60 hydrolase domain (fig. S2B). Conversely, the closest related protein in *E. gallinarum* showed only 67% sequence identity in the predicted catalytic domain. No direct orthologs were found in *E. faecalis*, and the two most similar conserved enzymes, *sagA*-like proteins A and B (SalA and SalB) (*23*), possessed distinct C-terminal domains. Structural modeling revealed that the secondary structural organization of putative SagA orthologs was similar to the crystal structure of the NlpC/p60 hydrolase domain from *E. faecium* SagA (fig. S4).

To determine whether these SagA-like enzymes are found in human patients treated with cancer immunotherapy, we performed sequence alignments using the NlpC/p60 domain of *E. faecium* Com15 SagA on clinical metagenomic data obtained from patients before checkpoint blockade treatment (*10*–*14*). However, we were unable to find SagA orthologs within the datasets, likely owing to low genome coverage of *E. faecium* in the samples (9.34%) (table S3). Therefore, we evaluated whether SagA-like enzymes are conserved across all sequenced strains of active enterococci, which include both clinical and environmental isolates. Query of all RefSeq database genomes from *E. faecium*, *E. durans*, *E. hirae*, and *E. mundtii* showed SagA orthologs to be almost universally present and conserved (Fig. 2B and table S4). We further confirmed that SagA orthologs are found in humans through alignment of the SagA NlpC/p60 domain with genomes from the Human Microbiome Project (*24*) (Fig. 2C and table S5). As expected, we saw nearly 100% conservation of the sequence within *E. faecium* isolates as well as a complete lack of SagA orthologs within *E. faecalis*. Together, these data indicate that SagA orthologs are present and conserved across active *Enterococcus* species found in the human microbiome.

To directly detect the expression of SagA orthologs in enterococci, we performed Western blotting on both the secreted and cell-associated protein fractions from each species. Using antiserum raised against *E. faecium* SagA (*19*), strong signals were observed in the supernatant fractions of *E. faecium*, *E. durans*, *E. hirae*, and *E. mundtii* but not *E. faecalis* or *E. gallinarum* (Fig. 2D and fig. S2C), indicating that SagA-like proteins are both highly expressed and secreted by these species. The signal was confirmed to be SagA dependent using a strain of *E. faecalis* OG1RF engineered with a chromosomal *sagA* insertion (Efs-*sagA*). As expected, all tested strains of *E. faecium* showed similar expression and secretion patterns for SagA (fig. S2D). To determine whether these SagA orthologs were functional, we analyzed the hydrolytic activity of the purified recombinant proteins on peptidoglycan in vitro (Fig. 2E and fig. S5). Proteins from *E. durans*, *E. hirae*, and *E. mundtii* showed D,L-endopeptidase activity against a model cross-linked peptidoglycan fragment similar to *E. faecium* SagA to produce the muropeptide GlcNAc-muramyl dipeptide (GMDP) (Fig. 2E and fig. S6). The hydrolytic activity of SagA was confirmed using a mutant construct lacking the cysteine active site residue, which was conserved in all orthologs (fig. S7). *E. faecalis* SalB did not hydrolyze peptidoglycan at detectable levels in our assay, whereas SalA cleaved the cross-linked fragment in the cross-bridge region rather than the peptide stem (Fig. 2E

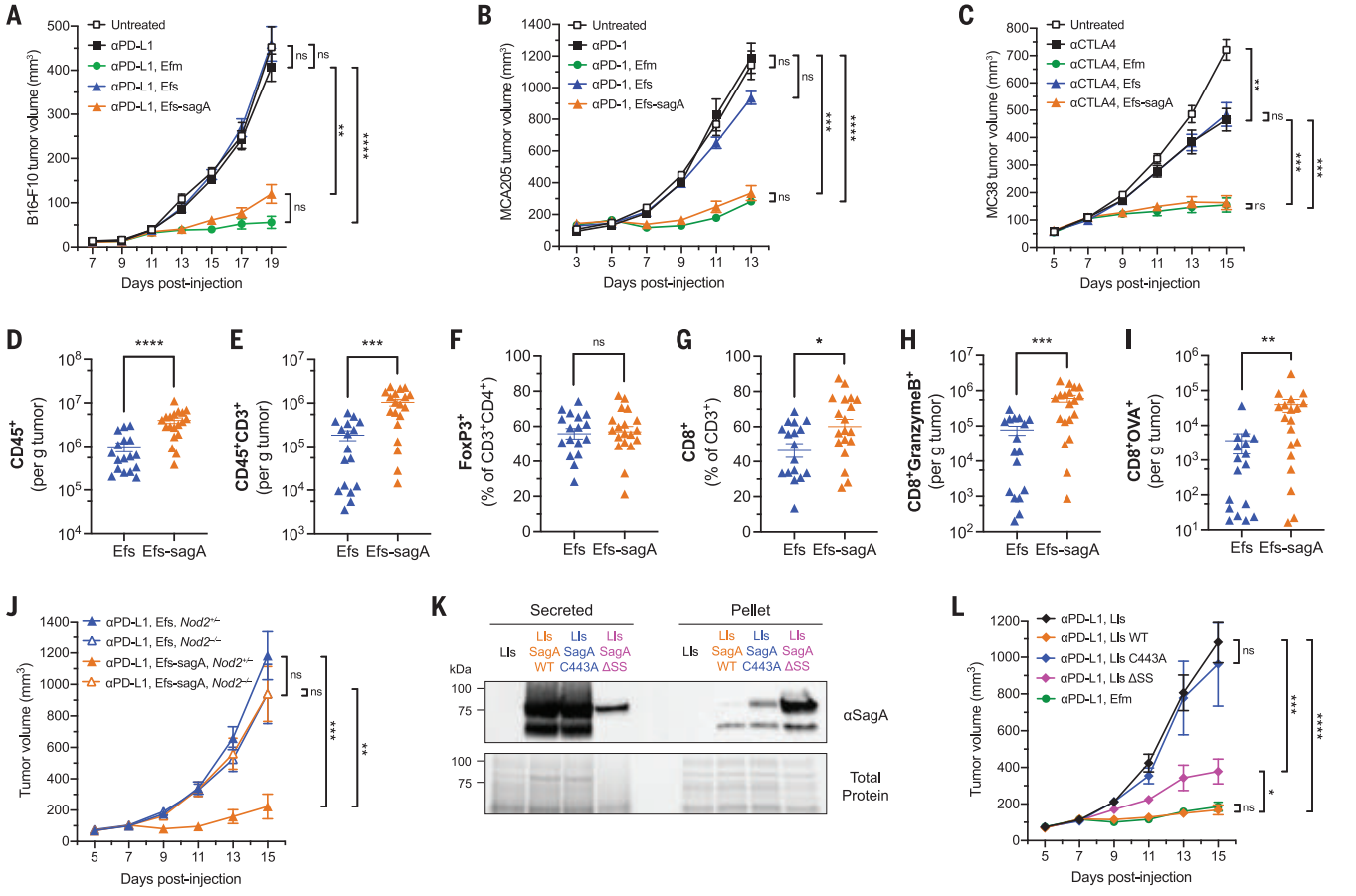


Fig. 3. SagA improves checkpoint inhibitor immunotherapy and elicits an adaptive immune response. (**A**) B16-F10 tumor growth in antibiotic-treated mice that were supplemented with the indicated enterococci and treated with anti–PD-L1 starting on day 9. *n* = 8 mice per group. (**B**) MCA205 tumor growth in antibiotic-treated mice that were supplemented with the indicated enterococci and treated with anti–PD-1 starting on day 5. *n* = 8 mice per group. (**C**) MC38 tumor growth in antibiotic-treated mice that were supplemented with the indicated enterococci and treated with anti–CTLA-4 starting on day 7. *n* = 8 mice per group. (**D** to **I**) Quantification of tumor infiltrating CD45$^+$ cells (D), total CD3$^+$ T cells (E), FoxP3$^+$ regulatory T cells (F), CD8$^+$ T cells (G), granzyme B$^+$ CD8$^+$ T cells (H), and OVA-specific CD8$^+$ T cells (I) from B16-OVA tumors in mice supplemented with *E. faecalis* or Efs-*sagA* harvested 5 days after the start of anti–PD-L1 treatment by flow cytometry. Data are pooled from two independent experiments with 7 to 10 mice per group per experiment; each symbol represents one mouse. (**J**) B16-F10 tumor growth in antibiotic-treated *Nod2*$^{+/-}$ or *Nod2*$^{-/-}$ mice that were supplemented with Efs or Efs-*sagA* and treated with anti–PD-L1 starting on day 7. *n* = 9 to 11 mice per group. (**K**) Western blot detection of ectopically expressed *E. faecium* Com15 SagA in secreted protein and cell pellet fractions harvested from overnight cultures of the indicated engineered *L. lactis* strains using antiserum raised against *E. faecium* Com15 SagA. Bottom panels show total protein loading. Numbers indicate estimated molecular weight (kDa). WT, wild type; ΔSS, signal sequence deletion. (**L**) B16-F10 tumor growth in antibiotic-treated mice that were supplemented with the indicated *L. lactis* strains and treated with anti–PD-L1 starting on day 7. *n* = 9 to 11 mice per group. Data for (A) to (C) and (K) and (L) represent mean ± SEM. and were analyzed using a mixed-effects model with Tukey's multiple comparisons post hoc test. Data for (D) to (I) represent mean ± SEM. and were analyzed by the Mann-Whitney *U* test. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001; ns, not significant.

and fig. S8). These results show that immunotherapy-active enterococci possess similar peptidoglycan composition and remodeling activity.

We then investigated whether SagA was sufficient to enhance the efficacy of anti–PD-L1 checkpoint inhibitor immunotherapy. Because *sagA* is an essential gene in *E. faecium* (*25*), we compared the inactive, parental *E. faecalis* OG1RF strain with the engineered, SagA-expressing strain *E. faecalis–sagA* (Fig. 2D). The peptidoglycan profile of *E. faecalis–sagA* showed changes consistent with increased D, L-endopeptidase activity (fig. S9A), indicative of active SagA expression (Fig. 2D). Animals that were treated with antibiotics and then orally supplemented with *E. faecalis–sagA* showed a significant decrease in B16-F10 tumor growth upon anti–PD-L1 therapy relative to those treated with the parental *E. faecalis* strain (Fig. 3A). This antitumor activity was similar to that of the phenotype observed in *E. faecium*–supplemented animals. Furthermore, fecal CFU analysis showed that *E. faecalis–sagA* exhibited a similar bacterial load to that of the parental *E. faecalis* strain and *E. faecium* (fig. S9B). To ensure that antitumor effects were not a result of overall changes in microbiome composition, we performed 16*S* rRNA analysis of fecal samples collected from our antibiotics-treated model system. Antibiotic pretreatment yielded a sharp decrease in read counts, with a maximal decrease observed at 2 weeks (fig. S10A). Supplementation with the tested strains led to a high bacterial load by day 3 (fig. S10A), with nearly 100% of the detected operational taxonomic units (OTUs) as *Enterococcus* (Fig. S10, B to E). Moreover, the microbiomes maintained ~5% abundance of *Enterococcus* 14 days after the start of supplementation, with an observed increase in spore-forming bacteria (fig. S10, D and E) likely resulting from incomplete clearance by antibiotics. As expected, beta diversity analysis

confirmed that the composition of the samples changed over time after antibiotic treatment and *Enterococcus* supplementation (fig. S10F). Permutational multivariate analysis of variance (ANOVA) testing of bacterial populations at day 14 showed no significant differences between treatment groups using both Bray-Curtis and weighted Unifrac principal coordinate analyses (PCoA) (fig. S10G). Moreover, OTU abundance quantification did not reveal other OTUs that were consistently correlated with biological activity (table S6), with most differentially prevalent OTUs found in only one of the two cages per condition.

We then tested the outcome of SagA expression on checkpoint inhibitor therapy in an animal model with an intact, complex microbiota. SPF C57BL/6 mice from Taconic Biosciences were chosen, as previous reports have found that these mice do not respond strongly to anti–PD-L1 therapy similar to germ-free

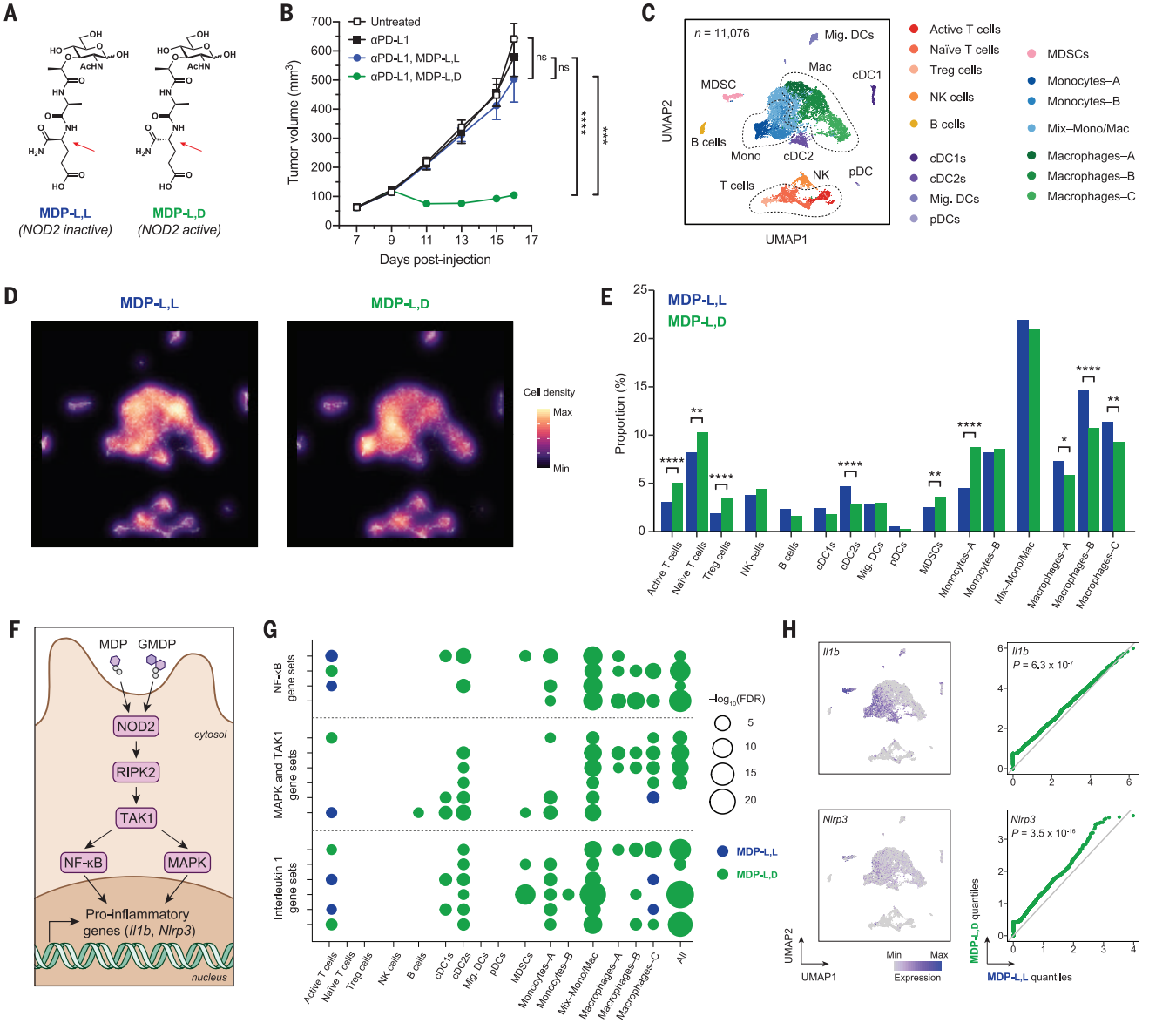


Fig. 4. Peptidoglycan fragment MDP enhances checkpoint blockade efficacy and generates a proinflammatory tumor microenvironment. (**A**) Chemical structures of the active (L,D) and inactive (L,L) diastereomers of MDP. Arrows indicate the single altered stereocenter. (**B**) B16-F10 tumor growth in antibiotic-treated, nonsupplemented mice treated with anti–PD-L1 and either MDP-$_{L,D}$ or MDP-$_{L,L}$ starting on day 9. n = 7 to 8 mice per group. (**C**) Uniform manifold approximation and projection (UMAP) plots of scRNA-seq data from CD45^{+} tumor-infiltrating cells after treatment with anti–PD-L1 and MDP diastereomers. T_{reg} cells, regulatory T cells; cDC1s/cDC2s, conventional type 1 or type 2 dendritic cells; Mig. DCs, migratory dendritic cells; pDCs, plasmacytoid dendritic cells; MDSCs, myeloid-derived suppressor cells. n = 11,076 total cells pooled from six animals per condition. (**D**) Density plot and (**E**) quantification of immune cell clusters. (**F**) Schematic of NOD2-dependent signaling of proinflammatory genes. (**G**) Bubble plot for enrichment of curated canonical pathway gene sets involving NF-κB, MAPK/TAK1, and interleukin-1 across cell clusters. (**H**) UMAP plots and paired quantile-quantile plots of NF-κB target genes *Il1b* and *Nlrp3*. For (B), data represent mean ± SEM and were analyzed by a mixed effects model with Tukey's multiple comparisons post hoc test. For (E), data represent absolute cell counts and were analyzed by Pearson's chi-square test for count data with Holm's correction for multiple comparisons. For (G), false discovery rates (FDRs) were obtained using model-based analysis of single transcriptomics. For (H), data were analyzed with a two-sided Wilcoxon rank-sum test. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001; ns, not significant.

animals (*9*, *13*). To maintain the resident nonresponsive microbiota, we omitted antibiotic treatment in this experiment and instead directly supplemented the mice with enterococci before B16-F10 implantation (fig. S11A). As seen in the antibiotic-pretreated model, animals supplemented with *E. faecium* or *E. faecalis–sagA* showed a significant decrease in tumor growth compared with treatment using either only the antibody or the antibody and parental *E. faecalis* (fig. S11B). *Enterococcus* CFU counts were much lower overall and did not significantly differ between the supplemented mice, suggesting that this effect did not require enteric domination by the newly administered, active enterococci (fig. S11C). Moreover, we observed no major changes in observed OTU counts, alpha diversity, or taxonomic composition upon *Enterococcus* administration throughout the experiment (fig. S11, D to F). On day 2 after the start of bacterial supplementation, no significant difference in microbiota composition was observed between samples by beta diversity analysis (fig. S11G). Permutational multivariant ANOVA testing of the weighted UniFrac PCoA also showed no difference between groups at day 14 (fig. S11H). Finally, differential OTU abundance analysis revealed no OTUs correlating with biological activity (table S7), suggesting that the observed anti-tumor effects were *Enterococcus*-driven.

We next investigated whether bacteria expressing SagA would also improve the efficacy of other checkpoint antibodies against different cancer cell types beyond anti–PD-L1 treatment of B16-F10 melanoma. Subcutaneous tumors were established in antibiotic-pretreated, enterococci-colonized animals with MCA205 fibrosarcoma or MC38 colorectal carcinoma cells and then treated with anti–PD-1 or anti–CTLA-4 immunotherapy, respectively. In both cases, we also observed a significant decrease in tumor growth when animals were cotreated with SagA-expressing enterococci and a checkpoint inhibitor (Fig. 3, B and C). Given the broad efficacy of these enterococci with different targeted therapies, we then asked whether this effect was mediated by an adaptive immune response. Using our antibiotic-pretreated model, we subcutaneously implanted B16-OVA tumor cells into animals colonized with either parental *E. faecalis* or *E. faecalis–sagA* and then treated with anti–PD-L1. Tumors were harvested the day after the third antibody treatment (5 days after the start of treatment), and tumor-infiltrating lymphocytes were quantified by flow cytometric analysis (fig. S12). Animals colonized with *E. faecalis–sagA* showed an overall increase in the absolute amount of intratumoral $CD45^+$ leukocytes as well as $CD3^+$ lymphocytes (Fig. 3, D and E). The composition of tumor-infiltrating $CD3^+$ lymphocytes showed an increase in the proportion of $CD8^+$ T cells but no change in $CD4^+FoxP3^+$ regulatory T cells (Fig. 3, F and G). Tumors contained higher amounts of $CD8^+$ T cells that expressed granzyme B (Fig. 3H), a marker for activated cytotoxic T lymphocytes (*26*). Tetramer staining also revealed a significant increase in the number of OVA-specific $CD8^+$ T cells (Fig. 3I), consistent with enhanced priming of a tumor antigen-specific immune response (*9*, *13*).

To characterize the microbial mechanism of immune activation, we first examined whether SagA contributed to bacterial dissemination, as *Enterococcus* translocation from the gut has been implicated during autoimmunity (*20*), chemotherapy treatment (*27*), and alcoholic hepatitis (*28*). CFU analysis of mesenteric lymph nodes and whole spleens of supplemented animals showed only low levels of live bacteria in proximal tissues (fig. S9C). The bacterial load of the mesenteric lymph nodes was independent of SagA expression, suggesting that SagA did not improve barrier transit for live bacteria in our study. We then turned our attention to peptidoglycan remodeling by SagA and the generation of muropeptides such as GMDP as a potential mechanism of action (*19*). We found that NOD2, a key pattern recognition receptor for muropeptides (*29*, *30*), was required for anti–PD-L1 antitumor activity in animals supplemented with *E. faecalis–sagA* (Fig. 3J). To evaluate the enzymatic activity and utility of SagA to improve checkpoint blockade, we investigated heterologous expression in probiotic bacteria. *Lactococcus lactis* (Lls) has been explored extensively as a live, oral probiotic to deliver bioactive proteins and enzymes (*31*). Therefore, we produced *L. lactis* strains that chromosomally expressed wild-type, catalytically inactive (C443A), or secretion-deficient (ΔSS) SagA. All three constructs effectively expressed SagA, and the wild-type and C443A mutant SagA constructs showed stronger signals in the secreted fraction, as expected (Fig. 3K). Antibiotic-pretreated animals were orally supplemented with these strains as well as parental *L. lactis* or *E. faecium* and were used to monitor B16-F10 responsiveness to anti–PD-L1. Animals supplemented with *L. lactis* expressing wild type SagA showed similar tumor growth inhibition to that of *E. faecium*–treated animals (Fig. 3L). *L. lactis* expressing catalytically inactive SagA (C443A) were not able to recapitulate the antitumor phenotype, indicating that the enzymatic activity of SagA is required to impede tumor growth. We found that the SagA secretion-deficient strain (ΔSS) did slow tumor growth, which suggests that the low amount of SagA secreted by this strain may have been sufficient to partially inhibit tumor growth. Alternatively, active, nonsecreted SagA may escape from *L. lactis* as a result of cell lysis in the gut.

Because SagA NlpC/p60 hydrolase activity was required, we asked whether synthetic muropeptide analogs of SagA enzymatic products such as muramyl dipeptide (MDP) could also elicit an improved response to checkpoint therapy. For these experiments, the NOD2-active MDP-$_{L,D}$ isomer or the inactive MDP-$_{L,L}$ diastereomer (Fig. 4A) were co-administered by intraperitoneal injection with anti–PD-L1. Animals that received active MDP-$_{L,D}$ along with anti–PD-L1 showed a significant antitumor effect that was not observed upon coadministration of the MDP-$_{L,L}$ negative control, which suggests that the NOD2-active muropeptide agonist was sufficient to improve checkpoint blockade (Fig. 4B). To better understand the mechanisms by which muropeptides augment checkpoint blockade, we profiled tumor-infiltrating leukocytes in the B16 melanoma model using single-cell–based RNA sequencing (scRNA-seq). After three treatments with anti–PD-L1 and MDP, $CD45^+$ cells from dissociated tumors were sorted, pooled, and sequenced (Fig. 4C and fig. S13). Tumors treated with the active MDP-$_{L,D}$ isomer showed a significant increase in the proportion of intratumoral T lymphocytes (Fig. 4, D and E), similar to our results using flow cytometry (Fig. 3F). The MDP-$_{L,D}$ sample also showed higher levels of the checkpoint genes *Ctla4* and *Pdcd1*, consistent with increased T cell receptor–mediated signaling (fig. S14A). Moreover, marker genes of cytotoxic T cell activity—including *Ifng*, *Gzmb*, *Ltb*, and *Prf1*—were enriched in MDP-$_{L,D}$–treated tumors (fig. S14B). We also observed significant shifts in myeloid cell populations after MDP-$_{L,D}$ administration (Fig. 4, D and E), with decreases in all macrophage clusters and an increase in a specific monocyte population characterized by *Cx3cr1* and *Nr4a1* expression (fig. S14C). NOD2 activation has previously been linked to a similar increase of $Cx3cr1^+$ monocytes with patrolling activity in circulation (*32*), and these results provide direct evidence that this monocyte subclass can accumulate in inflammatory microenvironments. Infrequent but detectable *Nod2* expression was observed across multiple myeloid populations as expected (*29*, *30*) (fig. S14D). Significant enrichment of hallmark gene sets (*33*) for inflammatory response and nuclear factor κB (NF-κB) signaling were found in the MDP-$_{L,D}$ dataset (table S8). Further uncurated gene set enrichment analysis revealed widespread changes in inflammatory, metabolic, and other innate immune pathways within myeloid cell clusters (tables S9 to S24). NOD2 signal transduction occurs through activation of transcription factor NF-κB and mitogen-activated protein kinase (MAPK) phosphorylation cascades by means of RIPK2 and TAK1 (Fig. 4F) (*29*, *30*). Accordingly, we found enrichment of multiple canonical gene sets involving these proteins, particularly within the monocyte and macrophage clusters that shifted upon muropeptide treatment (Fig. 4G). As further confirmation of increased NF-κB activation,

we found that NF-κB target transcripts *Il1b* and *Nlrp3* (*34*) were increased in the MDP-$_{L,D}$ sample along with IL-1β proinflammatory gene sets throughout the intratumoral myeloid compartment (Fig. 4, G and H).

Together, our data indicate that enterococci with distinctive NlpC/p60 peptidoglycan hydrolase activity can generate NOD2-active muropeptides and modulate the efficacy of checkpoint blockade immunotherapy in vivo (fig. S15). Although enriched bacteria in responding patients do not correlate well in regard to phylogeny (*35*), specific enterococci and other microbiota species with privileged cell wall composition and remodeling activity could provide functional indicators of therapeutic efficacy. Our results demonstrating the prevalence of conserved SagA-like enzymes throughout human-associated *Enterococcus* species indicate that the production of immune active muropeptides may be prevalent across human microbiomes. Beyond enterococci, the detection of potential NlpC/p60 orthologs in other genera such as *Lactobacillus* (Fig. 2C and table S5), which has also been identified in immunotherapy-responsive patients (*13*), may provide additional specific microbiota correlations to predict and improve patient outcomes. In addition, recent genetic analyses of *Bifidobacterium bifidium* strains that synergized with PD-1 blockade showed an enrichment of peptidoglycan biosynthetic genes (*36*). These observations suggest that peptidoglycan remodeling may be a broad mechanism for augmentation of immunotherapy efficacy, which requires additional functional studies of specific microbiota peptidoglycan remodeling factors. As peptidoglycan fragments can disseminate into circulation and prime systemic immune responses (*37*, *38*), the presence of NlpC/p60 hydrolases and NOD2-active muropeptides may be clinically relevant for predicting individual therapeutic responses. In addition, our results suggest that peptidoglycan remodeling enzymes may be used to reprogram probiotic bacteria as a novel therapeutic approach for enhancing the efficacy of checkpoint blockade inhibitors. Lastly, our findings corroborate a wealth of emerging evidence that NOD2 stimulation by MDP and its analogs can alter host immunity through multiple pathways, including direct activation of macrophages for tumor cell clearance (*39*), epigenetic reprogramming of monocytes (*40*, *41*), generation of conventional type 1 dendritic cells (*42*), and priming of dendritic cells for cross-presentation to $CD8^+$ T cells (*43*), which together have been implicated in trained immunity (*44*). Therefore, our work emphasizes how microbially produced or synthetic small molecules that can activate peptidoglycan pattern recognition receptors could be employed as next-generation adjuvants for immunotherapy.

REFERENCES AND NOTES

1. U. Sahin, Ö. Türeci, *Science* **359**, 1355–1360 (2018).
2. C. H. June, R. S. O'Connor, O. U. Kawalekar, S. Ghassemi, M. C. Milone, *Science* **359**, 1361–1365 (2018).
3. A. Ribas, J. D. Wolchok, *Science* **359**, 1350–1355 (2018).
4. E. M. Van Allen *et al*., *Science* **350**, 207–211 (2015).
5. M. Łuksza *et al*., *Nature* **551**, 517–520 (2017).
6. S. Spranger, R. Bao, T. F. Gajewski, *Nature* **523**, 231–235 (2015).
7. M. Binnewies *et al*., *Nat. Med.* **24**, 541–550 (2018).
8. M. Vétizou *et al*., *Science* **350**, 1079–1084 (2015).
9. A. Sivan *et al*., *Science* **350**, 1084–1089 (2015).
10. A. E. Frankel *et al*., *Neoplasia* **19**, 848–855 (2017).
11. V. Gopalakrishnan *et al*., *Science* **359**, 97–103 (2018).
12. B. Routy *et al*., *Science* **359**, 91–97 (2018).
13. V. Matson *et al*., *Science* **359**, 104–108 (2018).
14. B. A. Peters *et al*., *Genome Med.* **11**, 61 (2019).
15. F. Lebreton *et al*., *Cell* **169**, 849–861.e13 (2017).
16. H. Hanchi, W. Mottawea, K. Sebei, R. Hammami, *Front. Microbiol.* **9**, 1791 (2018).
17. K. J. Rangan *et al*., *Science* **353**, 1434–1437 (2016).
18. V. A. Pedicord *et al*., Exploiting a host-commensal interaction to promote intestinal barrier function and enteric pathogen tolerance. *Sci. Immunol.* **1**, eaai7732 (2016).
19. B. Kim *et al*., *eLife* **8**, e45343 (2019).
20. S. Manfredo Vieira *et al*., *Science* **359**, 1156–1161 (2018).
21. C. K. Stein-Thoeringer *et al*., *Science* **366**, 1143–1149 (2019).
22. A. Vermassen *et al*., *Front. Microbiol.* **10**, 331 (2019).
23. J. A. Mohamed, F. Teng, S. R. Nallapareddy, B. E. Murray, *J. Infect. Dis.* **193**, 231–240 (2006).
24. K. E. Nelson *et al*., *Science* **328**, 994–999 (2010).
25. F. Teng, M. Kawalec, G. M. Weinstock, W. Hryniewicz, B. E. Murray, *Infect. Immun.* **71**, 5033–5041 (2003).
26. J. Lieberman, *Nat. Rev. Immunol.* **3**, 361–370 (2003).
27. R. Daillère *et al*., *Immunity* **45**, 931–943 (2016).
28. Y. Duan *et al*., *Nature* **575**, 505–511 (2019).
29. D. J. Philpott, M. T. Sorbara, S. J. Robertson, K. Croitoru, S. E. Girardin, *Nat. Rev. Immunol.* **14**, 9–23 (2014).
30. R. Caruso, N. Warner, N. Inohara, G. Núñez, *Immunity* **41**, 898–908 (2014).
31. P. A. Bron, M. Kleerebezem, *Front. Microbiol.* **9**, 1821 (2018).
32. A.-J. Lessard *et al*., *Cell Rep.* **20**, 1830–1843 (2017).
33. A. Liberzon *et al*., *Cell Syst.* **1**, 417–425 (2015).
34. K. Taniguchi, M. Karin, *Nat. Rev. Immunol.* **18**, 309–324 (2018).
35. V. Gopalakrishnan, B. A. Helmink, C. N. Spencer, A. Reuben, J. A. Wargo, *Cancer Cell* **33**, 570–580 (2018).
36. S.-H. Lee *et al*., *Nat. Microbiol.* **6**, 277–288 (2021).
37. T. B. Clarke *et al*., *Nat. Med.* **16**, 228–231 (2010).
38. Z. Huang *et al*., *Nat. Microbiol.* **4**, 766–773 (2019).
39. I. J. Fidler, S. Sone, W. E. Fogler, Z. L. Barnes, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1680–1684 (1981).
40. J. Kleinnijenhuis *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17537–17542 (2012).
41. B. Priem *et al*., *Cell* **183**, 786–801 (2020).
42. D. Prescott *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10946–10957 (2020).
43. J. Asano *et al*., *J. Immunol.* **184**, 736–745 (2010).
44. M. G. Netea *et al*., *Nat. Rev. Immunol.* **20**, 375–388 (2020).
46. S. Kommineni *et al*., *Nature* **526**, 719–722 (2015).
47. O. De Henau *et al*., *Nature* **539**, 443–447 (2016).
48. D. J. McCarthy, K. R. Campbell, A. T. L. Lun, Q. F. Wills, *Bioinformatics* **33**, 1179–1186 (2017).
49. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, *Nat. Biotechnol.* **36**, 411–420 (2018).
50. T. Stuart *et al*., *Cell* **177**, 1888–1902.e21 (2019).
51. G. Finak *et al*., *Genome Biol.* **16**, 278 (2015).
52. A. Subramanian *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
53. D. Arndt *et al*., *Nucleic Acids Res.* **44** (W1), W16-21 (2016).
54. F. Madeira *et al*., *Nucleic Acids Res.* **47** (W1), W636–W641 (2019).
55. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, *Mol. Biol. Evol.* **32**, 268–274 (2015).
56. J. Trifinopoulos, L.-T. Nguyen, A. von Haeseler, B. Q. Minh, *Nucleic Acids Res.* **44** (W1), W232–W235 (2016).
57. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, *Mol. Biol. Evol.* **35**, 518–522 (2018).
58. I. Letunic, P. Bork, *Nucleic Acids Res.* **47** (W1), W256–W259 (2019).
59. X. Robert, P. Gouet, *Nucleic Acids Res.* **42** (W1), W320-4 (2014).
60. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat. Protoc.* **10**, 845–858 (2015).
61. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, *Bioinformatics* **25**, 1972–1973 (2009).
62. M. N. Price, P. S. Dehal, A. P. Arkin, *PLOS ONE* **5**, e9490 (2010).
63. B. Kim, J. Espinosa, H. C. Hang, *Methods Enzymol.* **638**, 109–127 (2020).
64. J. G. Caporaso *et al*., *ISME J.* **6**, 1621–1624 (2012).
65. R. C. Edgar, *Bioinformatics* **26**, 2460–2461 (2010).
66. R. C. Edgar, *bioRxiv* (2016), doi: 10.1101/074161.
67. J. R. Cole *et al*., *Nucleic Acids Res.* **42** (D1), D633–D642 (2014).
68. P. J. McMurdie, S. Holmes, *PLOS ONE* **8**, e61217 (2013).
69. M. I. Love, W. Huber, S. Anders, *Genome Biol.* **15**, 550 (2014).
70. Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
71. K. J. Leenhouts, J. Kok, G. Venema, *J. Bacteriol.* **173**, 4794–4798 (1991).
72. G. Simons, M. Nijhuis, W. M. de Vos, *J. Bacteriol.* **175**, 5168–5175 (1993).
73. Y. Sasaki, Y. Ito, T. Sasaki, *Appl. Environ. Microbiol.* **70**, 1858–1864 (2004).

ACKNOWLEDGMENTS

We thank S. Mazel, A. Keprova, M. Jaimes, D. Tran, and the Rockefeller Flow Cytometry Resource Center for assistance with flow cytometric experiments; G. Putzel and the Weill Cornell Microbiome Core for assistance with 16S rRNA sequencing experiments and analysis; C. Steckler and H. Molina of the Rockefeller Proteomics Resource Center for assistance with peptidoglycan analysis; K. Eckartt for assistance in homology modeling; B. Ostendorf for assistance with scRNA-seq data analysis; A. Griffin for assistance with bacterial genomic data preparation and analysis; and M. Huse, T. Merghoub, K. Cadwell, and D. Mucida for helpful discussions. **Funding:** This work was supported by the National Institutes of Health (1R01CA245292-01 to H.C.H.) and partly by the Melanoma Research Foundation (Career Development Award to M.E.G.) M.E.G. is a Hope Funds for Cancer Research Fellow supported by the Hope Funds for Cancer Research (HCFR-19-03-02). This work is also partially funded by an NIH research service award training grant (A1070084, J.E.) **Author contributions:** Conceptualization: M.E.G. and H.C.H.; Methodology: M.E.G., J.E., and H.C.H.; Investigation: M.E.G., J.E., and J.L.B.; Formal analysis: J.-D.L. and T.S.C.; Supervision: H.C.H.; Writing – original draft: M.E.G. and H.C.H.; Writing – review and editing: M.E.G., J.E., J.L.B., J.-D.L., T.S.C., J.K.J., G.R.F., and H.C.H.; Resources J.K.J. and G.R.F. **Competing interests:** M.E.G. and H.C.H. have filed a patent application (PCT/US2020/019038) for the commercial use of SagA-bacteria to improve checkpoint blockade immunotherapy. Rise Therapeutics (J.K.J. and G.R.F.) has licensed the patent to develop immunological-based biologics. **Data and materials availability:** All 16S rRNA and scRNA-seq sequencing data are available via BioProject PRJNA749064. All other data are available in the main text and supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/1040/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S24
References (*45*–*74*)

20 May 2020; accepted 27 July 2021
10.1126/science.abc9113

RNA

Geometric deep learning of RNA structure

Raphael J. L. Townshend[1]†‡, Stephan Eismann[1,2]†, Andrew M. Watkins[3]†, Ramya Rangan[3,4], Maria Karelina[1,4], Rhiju Das[3,5]*, Ron O. Dror[1,6,7,8]*

RNA molecules adopt three-dimensional structures that are critical to their function and of interest in drug discovery. Few RNA structures are known, however, and predicting them computationally has proven challenging. We introduce a machine learning approach that enables identification of accurate structural models without assumptions about their defining characteristics, despite being trained with only 18 known RNA structures. The resulting scoring function, the Atomic Rotationally Equivariant Scorer (ARES), substantially outperforms previous methods and consistently produces the best results in community-wide blind RNA structure prediction challenges. By learning effectively even from a small amount of data, our approach overcomes a major limitation of standard deep neural networks. Because it uses only atomic coordinates as inputs and incorporates no RNA-specific information, this approach is applicable to diverse problems in structural biology, chemistry, materials science, and beyond.

RNA molecules, like proteins, fold into well-defined three-dimensional (3D) structures to perform a wide range of cellular functions, such as catalyzing reactions, regulating gene expression, modulating innate immunity, and sensing small molecules (*1*). Knowledge of these structures is extremely important for understanding the mechanisms of RNA function, designing synthetic RNAs, and discovering RNA-targeted drugs (*2*, *3*). Our knowledge of RNA structure lags far behind that of protein structure: The fraction of the human genome transcribed to RNA is ~30 times as large as that coding for proteins (*4*), but the number of available RNA structures is <1% of that for proteins (*5*). Computational prediction of RNA 3D structure is thus of substantial interest (*6*).

Despite decades of intense effort, predicting the 3D structure of RNAs remains a grand challenge, having proven more difficult than prediction of protein structure. For proteins, state-of-the-art prediction methods leverage sequences or structures of related proteins (*7*–*9*). Such methods succeed much less frequently for RNA, both because template structures of closely related RNAs are available far less often and because sequence coevolution information provides less information about tertiary contacts in RNAs (*10*). Moreover, designing a scoring function that reliably distinguishes accurate structural models of RNA from less accurate ones has proven difficult, because the characteristics of energetically favorable RNA structures are not sufficiently well understood.

This problem raises the question of whether an algorithm could learn from known RNA structures to assess the accuracy of structural models of unrelated RNAs. Such a machine learning task poses two major challenges: (i) avoiding assumptions about which structural characteristics might distinguish accurate models from less accurate ones, and (ii) learning from the limited number of RNA structures that have been determined experimentally. Deep learning methods that do not require predefined features have led to notable recent advances in many fields, but their success has largely been restricted to domains where data are plentiful (*11*).

We designed a neural network, the Atomic Rotationally Equivariant Scorer (ARES), to address these challenges (Fig. 1). Given a structural model, specified by the 3D coordinates and chemical element type of each atom, ARES predicts the model's root mean square deviation (RMSD) from the unknown true structure. ARES is a deep neural network: It consists of many processing layers, with each layer's outputs serving as the next layer's inputs (*11*). This network has a distinctive architecture that enables it to learn directly from 3D structures and to learn effectively given a very small amount of experimental data.

ARES does not incorporate any assumptions about which features of a structural model are relevant to assessing its accuracy. For example, ARES has no preconceived notion of double helices, base pairs, nucleotides, or hydrogen bonds. The approach behind ARES is not at all specific to RNA and is thus applicable to any type of molecular system.

The initial layers of the ARES network are designed to recognize structural motifs, the identities of which are learned during the training process rather than specified in advance. Each of these layers computes several features for each atom based on the geometric arrangement of surrounding atoms and the features computed by the previous layer. The first layer's only inputs are the 3D coordinates and chemical element type of each atom.

The architecture of these initial network layers recognizes that instances of a given structural motif are typically oriented and positioned differently from one another and that coarser-scale motifs (e.g., helices) often comprise particular arrangements of finer-scale motifs (e.g., base pairs). Each layer is rotationally and translationally equivariant—that is, rotation or translation of its input leads to a corresponding transformation of its output (*12*). This property captures the invariance of physics to rotation or translation of the frame of reference but ensures that orientation and position of an identified motif are passed on to the network's next layer, which can use this information to recognize coarser-scale motifs. The design of these layers builds on recently developed machine learning techniques that capture rotational as well as translational symmetries (*13*–*15*)—particularly tensor field networks (*16*) and the PAUL method (*17*).

Whereas the initial layers of ARES gather information locally, the remaining layers aggregate information across all atoms. This combination allows ARES to predict a global property (in this case, the accuracy of the structural model) while capturing local structural motifs and interatomic interactions in detail.

To train ARES, we used 18 RNA molecules for which experimentally determined structures were published between 1994 and 2006 (*18*). We generated 1000 structural models of each RNA with the Rosetta FARFAR2 sampling method (*19*), without making any use of the known structure. We then optimized the parameters of the ARES neural network such that its output matches as closely as possible the RMSD of each model from the corresponding structure.

To assess ARES's ability to identify accurate structural models of previously unseen RNAs, we used a benchmark consisting of all RNAs that were included in the RNA-Puzzles structure prediction challenge and for which experimentally determined structures were published between 2010 and 2017 (*20*). For each of these RNAs, we generated at least 1500 structural models using FARFAR2 (*12*). To ensure that some models were near native (i.e., within a 2-Å RMSD of the experimentally determined native structure), we included energetic restraints to the native structure's coordinates when generating 1% of the models for each RNA (*12*). We used the trained ARES network to produce a score for each model (i.e., the predicted RMSD of each model from the native structure). We also scored each model using three state-of-the-art scoring functions: the most recent (2020) version of Rosetta

[1]Department of Computer Science, Stanford University, Stanford, CA, USA. [2]Department of Applied Physics, Stanford University, Stanford, CA, USA. [3]Department of Biochemistry, Stanford University, Stanford, CA, USA. [4]Biophysics Program, Stanford University, Stanford, CA, USA. [5]Department of Physics, Stanford University, Stanford, CA, USA. [6]Department of Structural Biology, Stanford University, Stanford, CA, USA. [7]Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA. [8]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA.
†These authors contributed equally to this work.
‡Present address: Atomic AI, Menlo Park, CA, USA.
*Corresponding author. Email: ron.dror@stanford.edu (R.O.D.); rhiju@stanford.edu (R.D.)

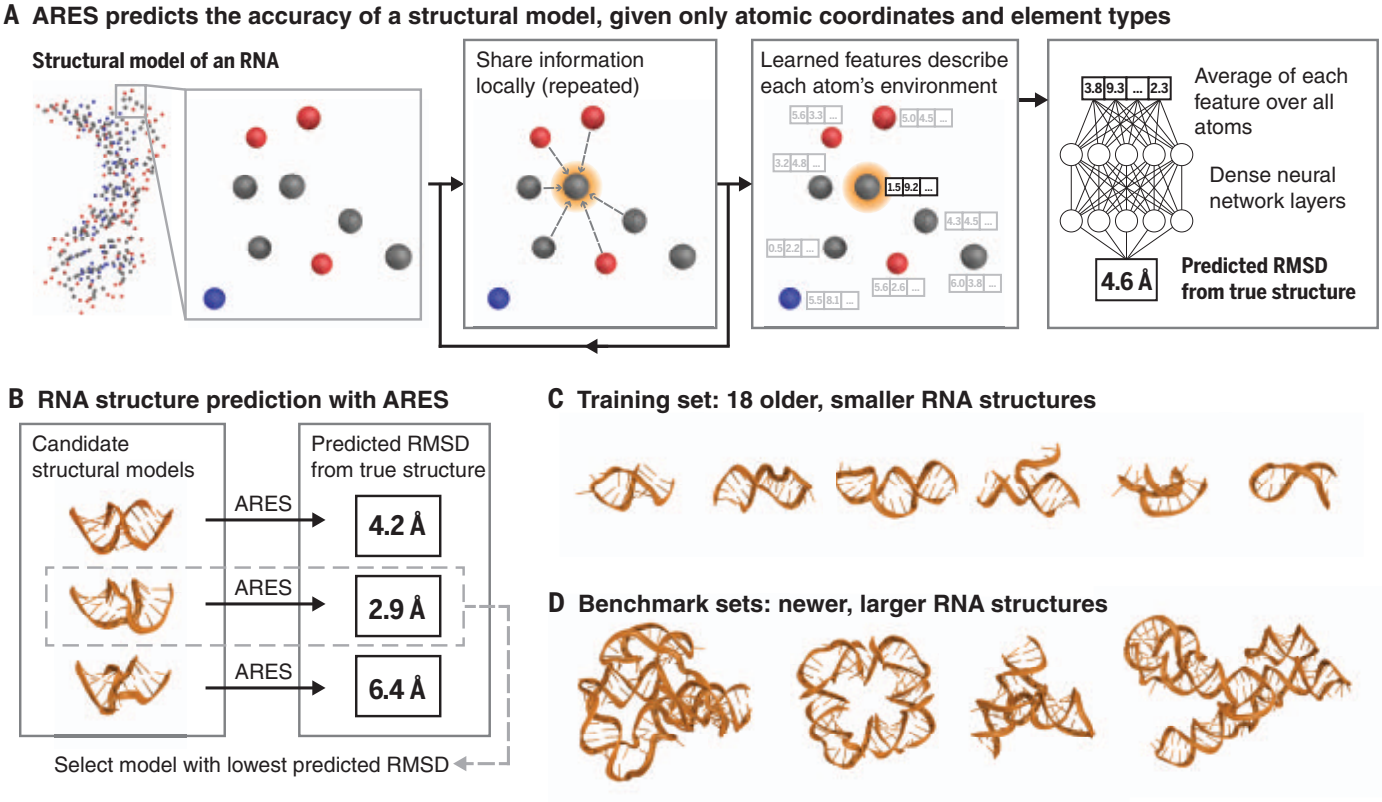


Fig. 1. The ARES network. (**A**) ARES takes as input a structural model, specified by each atom's element type and 3D coordinates. Atom features are repeatedly updated based on the features of nearby atoms. This process results in a set of features encoding each atom's environment. Each feature is then averaged across all atoms, and the resulting averages are fed into additional neural network layers, which output the predicted RMSD of the structural model from the true structure of the RNA molecule. Figure S1 illustrates the ARES architecture in more detail. (**B**) To perform structure prediction, we use ARES to score candidate structural models (e.g., those generated by the FARFAR2 sampling software), selecting the models that ARES predicts to be most accurate (i.e., lowest RMSD). (**C**) ARES is trained using 18 RNA structures solved before 2007. (**D**) We benchmark ARES using more recently solved RNA structures, most of which are much larger than any of those used for training. Representative examples of structures used for training and benchmarking are shown in this figure, with the remainder in fig. S2.

(*19*), Ribonucleic Acids Statistical Potential (RASP) (*21*), and 3dRNAscore (*22*).

ARES substantially outperforms the other three scoring functions on this first benchmark (Fig. 2, A to C, and figs. S3 and S4). The single best-scoring structural model is near native (<2 Å RMSD) for 62% of the benchmark RNAs when using ARES, compared with 43, 33, and 5% for Rosetta, RASP, and 3dRNAscore, respectively. The 10 best-scoring models include at least one near-native model for 81% of the benchmark RNAs when using ARES, compared with 48, 48, and 33% for Rosetta, RASP, and 3dRNAscore, respectively. Each of the best-scoring near-native models was generated with energetic restraints to the native structure.

Current methods for sampling candidate structural models often fail to generate near-native models in a reasonable amount of computation time. We therefore compiled a second benchmark that includes no near-native models. When predicting RNA structure, experts can often find some known structures that can be used as local templates, or other published experimental data that provide information on local tertiary structure—but we ignored all such prior information when generating candidate models, to simulate a difficult modeling scenario. We selected 16 structurally diverse RNAs, all substantially different from any of those used to train ARES or those in our previous benchmark, and each including one or more of the following hallmarks of structural complexity: ligand binding sites, multiway junctions, and tertiary contacts. We scored all models using ARES as well as six other scoring functions that have been used widely over the past 14 years.

On this second benchmark, ARES again outperforms all the other scoring functions (Fig. 2D and fig. S5). The median RMSD across RNAs of the best-scoring structural model is significantly lower for ARES than for any other scoring function. The same is true when considering the most accurate of the 10 best-scoring structural models for each RNA.

Next, we turned to blind prediction of 3D RNA structure, participating in four rounds of RNA-Puzzles, a long-running community-wide challenge in which newly determined experimental structures are held in confidence until all participants have submitted their structural predictions (*20*). For each RNA molecule, we generated candidate structural models using the FARFAR2 sampling protocol and then selected among these models using the ARES scoring function. We describe the exact inputs to FARFAR2 in table S1. No template information was available for one of these structure prediction challenges, an intricate adenoviral RNA. Limited template information was available for the other three, which are distinct T-box–tRNA complexes (fig. S6 and table S1).

In every case, this procedure yielded the most accurate model submitted by any participant, as measured by both RMSD and deformation index (Fig. 3, fig. S8, and table S2). For each RNA, competing submissions were produced by at least nine other methods, including methods that used the same sets of candidate FARFAR2 structural models but selected among them using the judgment of human experts in

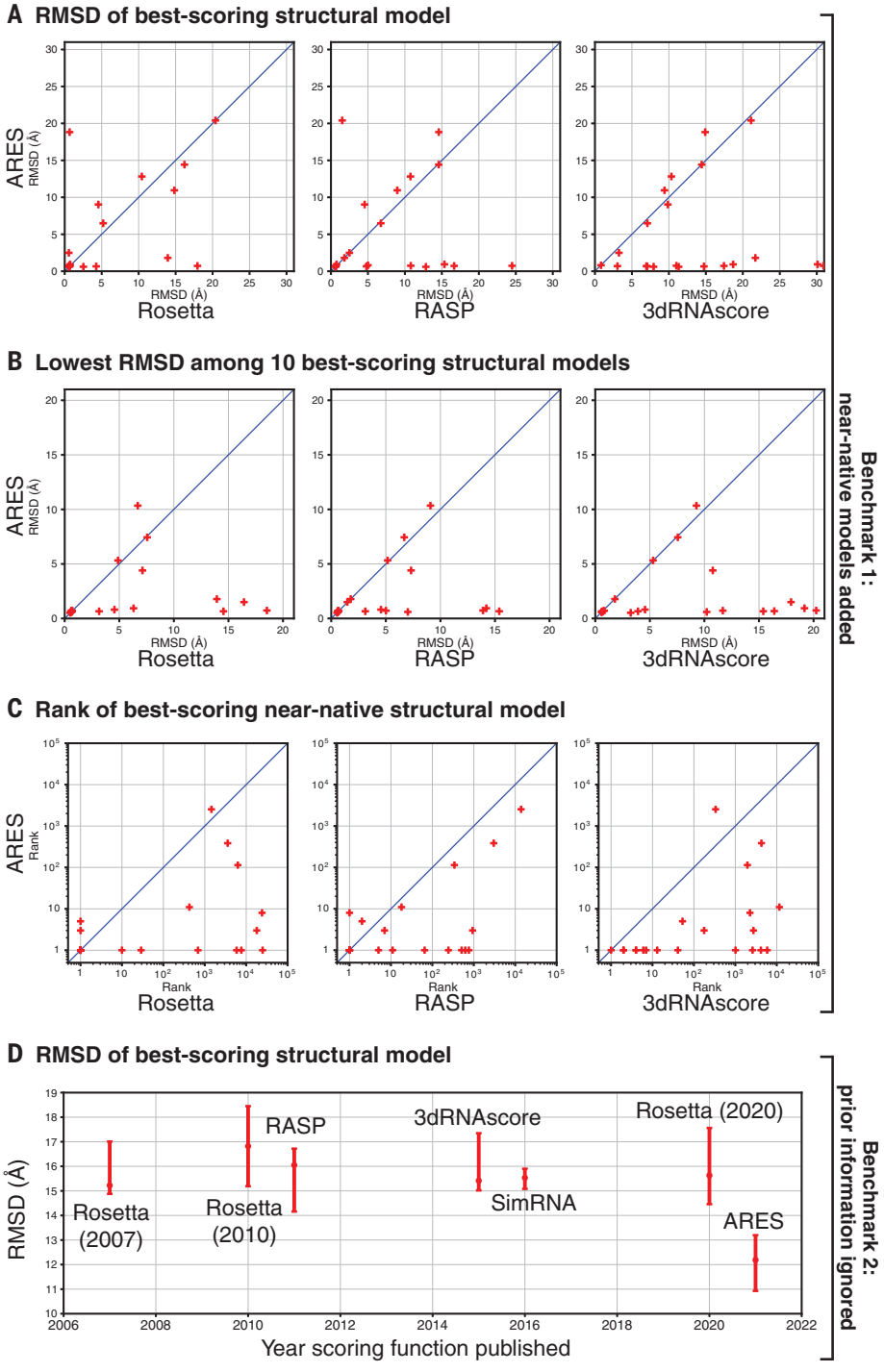


Fig. 2. ARES substantially outperforms previous scoring functions at identifying accurate structural models. (**A**) Given a large set of candidate structural models for each RNA in benchmark 1—which includes some models restrained to be close to the experimentally determined (native) structure—we rank the models using ARES and three leading scoring functions. The model scored as best by ARES is usually more accurate (as assessed by RMSD from the native structure) than the model scored as best by the other scoring functions. Each cross corresponds to one RNA. "Rosetta" indicates the most recent (2020) version of the Rosetta scoring function. (**B**) When using ARES, the 10 best-scoring structural models for each RNA in benchmark 1 include an accurate model more frequently than when using the other scoring functions. (**C**) For each RNA in benchmark 1, we determine the rank of the best-scoring near-native structural model—that is, how far down the ranked list we need to go to include one near-native structural model (RMSD < 2 Å). This rank is usually lower (better) for ARES than for the other scoring functions. Across the RNAs, the mean rank of the best-scoring near-native model is 3.6 for ARES, compared with 73.0, 26.4, and 127.7 for Rosetta, RASP, and 3dRNAscore, respectively (geometric means). (**D**) For each of the 16 RNAs in benchmark 2—for which all structural models were generated without using any template structures or other experimental data that could provide information on local tertiary structure—we determine the RMSD of the model scored as best by each of seven scoring functions. For each scoring function, we plot the median across RNAs, with a 95% confidence interval determined by bootstrapping (*12*). ARES significantly outperforms each of the other scoring functions [*P* values 0.001 to 0.016 (*12*)]. Of the other scoring functions, none significantly outperforms any other [*P* values 0.24 to 0.66].

the Das lab or the Rosetta (2020) scoring function. We also found that the ARES scoring function outperforms a variety of other scoring functions applied to the same sets of candidate models, including a recent machine learning approach based on a convolutional neural network (*23*) (table S3).

Analysis of the trained ARES network indicates that it has spontaneously discovered certain fundamental characteristics of RNA structure. For example, ARES correctly predicts the optimal distance between the two strands in a double helix—i.e., the distance that allows for ideal base pairing (Fig. 4A). In addition, the high-level features ARES extracts from a set of RNA structures reflect the extent of hydrogen bonding and Watson-Crick base pairing in each structure (Fig. 4B), even though we never informed ARES that hydrogen bonding and base pairing are key drivers of RNA structure formation. We also observe that ARES is able to accurately identify complex tertiary structure elements, including ones that are not represented in the training dataset (figs. S7 and S9).

Several deep learning techniques have recently been applied to problems in structural biology as well as quantum chemistry, leading to substantial advances in protein structure prediction and other areas (*8*, *24*–*30*). ARES, however, tackles a particularly challenging geometric learning problem, in that it (i) learns entirely from atomic structure, using no other information (e.g., sequences of related RNAs or

Fig. 3. ARES achieves state-of-the-art results in blind RNA structure prediction. (**A**) We submitted structural models, which ARES selected from sets of candidates generated by FARFAR2, to four recent rounds of the RNA-Puzzles blind structure prediction challenge: RNA A (the adenovirus VA-I RNA), RNA B (the *Geobacillus kaustophilus* T-box discriminator–$tRNA^{Gly}$), RNA C (the *Bacillus subtilis* T-box–$tRNA^{Gly}$), and RNA D (the *Nocardia farcinic* T-box–$tRNA^{Ile}$), whose structures are now in the Protein Data Bank with IDs 6OL3, 6PMO, 6POM, and 6UFM, respectively. For all four RNAs, ARES produced the most accurate structural model of any method. The dash indicates no submission. (**B**) For RNA A, the adenovirus VA-I RNA, ARES selected a structural model (blue) with a 4.8-Å RMSD to the experimentally determined structure (green), which was not available at prediction time. (**C**) For RNA A, the most accurate structural model produced by any another method (orange; produced by Rosetta) had an RMSD of 7.7 Å. ARES predicts the 3D geometry of the hinge motif at lower left much more accurately than Rosetta (fig. S7). Figure S8 illustrates results for the other RNAs.

A Blind prediction accuracy (RMSD, Å)

Method	RNA			
	A	B	C	D
ARES	**4.8**	**12.5**	**9.5**	**14.5**
Adamiak	9.8	18.7	19.1	18.2
Bujnicki	9.8	14.0	15.6	20.0
Chen	11.0	18.1	11.7	32.8
Ding	19.1	17.4	—	34.3
Das (Human)	13.6	13.3	10.1	28.8
iFoldRNA	10.3	23.5	53.3	22.4
RNAComposer	10.2	19.0	14.1	19.6
Rosetta	7.7	14.3	10.1	22.2
SimRNA	13.7	16.2	42.2	22.2
Xiao	15.4	20.6	27.2	29.4

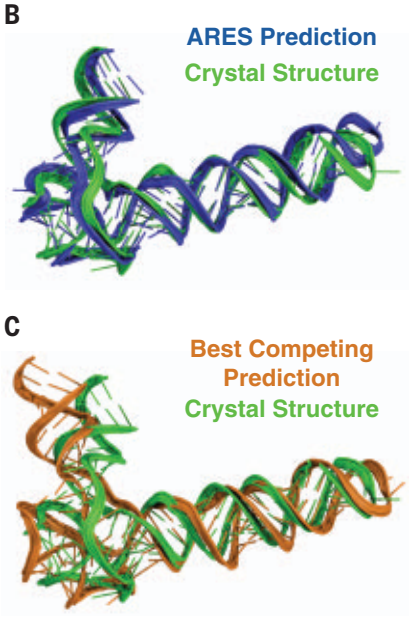
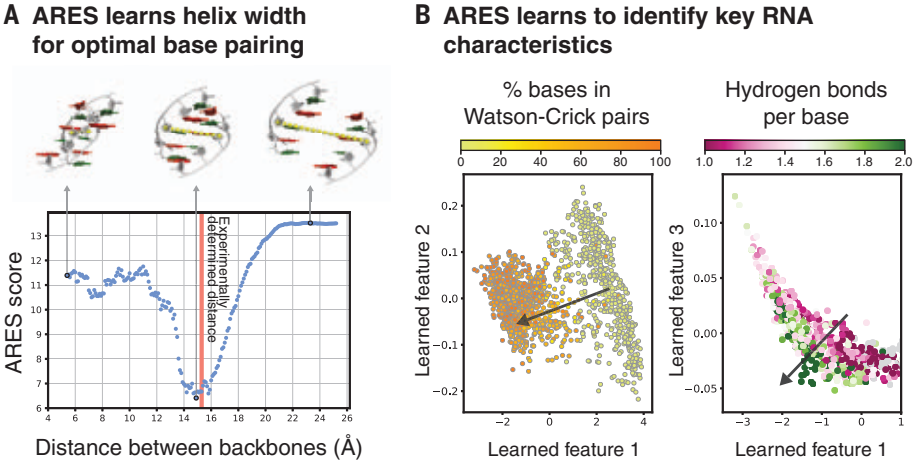


Fig. 4. ARES learns to identify key characteristics of RNA structure that are not specified in advance. (**A**) As the distance between two complementary strands of an RNA double helix is varied, ARES assigns the best scores when the distance closely approximates the experimentally observed distance (red vertical line). The distance is measured between the C-4′ atoms of the central base pair (yellow dotted lines). (**B**) ARES's learned features separate RNA structures according to the fraction of bases that form Watson-Crick pairs (left) and the average number of hydrogen bonds per base (right). The arrow in each plot indicates the direction of separation. Learned features 1, 2, and 3 are the first, second, and third principal components, respectively, of the activation values of the 256 nodes in ARES's penultimate layer across 1576 RNA structures. Each dot corresponds to one of these structures (*12*).



proteins), and (ii) makes no assumptions about what structural features might be important, taking inputs specified simply as atomic coordinates and without even being provided basic information such as the fact that RNAs comprise chains of nucleotides.

ARES's performance is particularly notable given that all of the RNAs used for blind structure prediction (Fig. 3), and most of those used for systematic benchmarking (Fig. 2), are larger and more complex than those used to train ARES (Fig. 1). RNAs in the training set comprise 17 to 47 nucleotides (median: 26), whereas RNAs in the blind structure prediction challenges comprise 112 to 230 nucleotides (median: 152.5), and RNAs in the benchmark sets comprise 27 to 188 nucleotides (median: 75, with 31 of 37 RNAs comprising more nucleotides than any RNA in the training set).

A limitation of the current study is its reliance on a previously developed sampling method to generate candidate structural models. Future work might use ARES to guide sampling, so as to increase the accuracy of the best candidate models. ARES might be improved further by incorporating other types of experimental data, including low-resolution cryogenic electron microscopy and chemical mapping data (*31*).

ARES's ability to outperform the previous state of the art despite using only a small number of structures for training suggests that similar neural networks could lead to substantial advances in other areas involving 3D molecular structure, where data are often limited and expensive to collect. In addition to structure prediction, examples might include molecular design (both for macromolecules such as proteins or nucleic acids and for small-molecule drugs), estimation of electromagnetic properties of nanoparticle semiconductors, and prediction of mechanical properties of alloys and other materials.

REFERENCES AND NOTES

1. T. R. Cech, J. A. Steitz, *Cell* **157**, 77–94 (2014).
2. K. D. Warner, C. E. Hajdin, K. M. Weeks, *Nat. Rev. Drug Discov.* **17**, 547–558 (2018).
3. A. Churkin *et al.*, *Brief. Bioinform.* **19**, 350–358 (2018).
4. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
5. H. M. Berman *et al.*, *Nucleic Acids Res.* **28**, 235–242 (2000).
6. S. Jain, D. C. Richardson, J. S. Richardson, *Methods Enzymol.* **558**, 181–212 (2015).
7. D. S. Marks *et al.*, *PLOS ONE* **6**, e28766 (2011).

8. A. W. Senior *et al*., . *Nature* **577**, 706–710 (2020).
9. H. Kamisetty, S. Ovchinnikov, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15674–15679 (2013).
10. F. Pucci, M. B. Zerihun, E. K. Peter, A. Schug, *RNA* **26**, 794–802 (2020).
11. Y. LeCun, Y. Bengio, G. Hinton, *Nature* **521**, 436–444 (2015).
12. Materials and methods are available as supplementary materials.
13. D. E. Worrall, S. J. Garbin, D. Turmukhambetov, G. J. Brostow, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2017), pp. 7168–7177.
14. B. Anderson, T. S. Hy, R. Kondor, in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, H. Wallach *et al*., Eds. (Curran Associates, 2019), pp. 14537–14546.
15. M. Weiler, M. Geiger, M. Welling, W. Boomsma, T. S. Cohen, in *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, S. Bengio *et al*., Eds. (Curran Associates, 2018), pp. 10381–10392.
16. N. Thomas *et al*., arXiv:1802.08219 [cs.LG] (2018).
17. S. Eismann *et al*., *Proteins* **89**, 493–501 (2021).
18. R. Das, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14664–14669 (2007).
19. A. M. Watkins, R. Rangan, R. Das, *Structure* **28**, 963–976.e6 (2020).
20. Z. Miao *et al*., *RNA* **26**, 982–995 (2020).
21. E. Capriotti, T. Norambuena, M. A. Marti-Renom, F. Melo, *Bioinformatics* **27**, 1086–1093 (2011).
22. J. Wang, Y. Zhao, C. Zhu, Y. Xiao, *Nucleic Acids Res.* **43**, e63 (2015).
23. J. Li *et al*., *PLOS Comput. Biol.* **14**, e1006514 (2018).
24. J. Behler, M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).
25. J. Xu, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16856–16865 (2019).
26. J. S. Smith, O. Isayev, A. E. Roitberg, *Chem. Sci.* **8**, 3192–3203 (2017).
27. M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, D. R. Koes, *J. Chem. Inf. Model.* **57**, 942–957 (2017).
28. K. Xu, Z. Wang, J. Shi, H. Li, Q. C. Zhang, in *Proceedings of the AAAI Conference on Artificial Intelligence* (AAAI Press, 2019), pp. 1230–1237.
29. F. Noé, S. Olsson, J. Köhler, H. Wu, *Science* **365**, eaaw1147 (2019).
30. M. AlQuraishi, *Cell Syst.* **8**, 292–301.e3 (2019).
31. K. M. Kutchko, A. Laederach, *WIREs RNA* **8**, e1374 (2017).
32. R. J. L. Townshend *et al*., Training code for ARES neural network, Version 1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.5088971.
33. R. J. L. Townshend *et al*., ARES-specific adaptation of E3NN, Version 1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.5090151.
34. R. J. L. Townshend *et al*., Auxiliary code related to the publication "Geometric Deep Learning of RNA Structure," Version 1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.5090157.
35. R. J. L. Townshend *et al*., Structural data used to train, test, and characterize a new geometric deep learning RNA scoring function, Stanford Digital Repository (2020); https://doi.org/10.25740/bn398fc4306.
36. A. M. Watkins *et al*., Structural data used to test a new geometric deep learning RNA scoring function emulating fully de novo modeling conditions, Stanford Digital Repository (2021); https://doi.org/10.25740/sq987cc0358.

ACKNOWLEDGMENTS

The authors thank R. Altman, R. Betz, T. Dao, H. Hanley, S. Hollingsworth, M. Jagota, B. Jing, Y. Laloudakis, N. Latorraca, J. Paggi, A. Powers, P. Suriana, and N. Thomas for discussions and advice. **Funding:** Funding was provided by National Science Foundation Graduate Research Fellowships (R.J.L.T. and R.R.); the U.S. Department of Energy, Office of Science Graduate Student Research program (R.J.L.T.); a Stanford Bio-X Bowes Fellowship (S.E.); the Army Research Office Multidisciplinary University Research Initiative program (R.D.); the U.S. Department of Energy, Office of Science, Scientific Discovery through Advanced Computing (SciDAC) program (R.O.D.); Intel (R.O.D.); a Stanford Bio-X seed grant (R.D. and R.O.D.); and National Institutes of Health grants R21CA219847 (R.D.) and R35GM122579 (R.D.).

Author contributions: R.J.L.T., A.M.W., R.D., and R.O.D. designed the research. S.E. formulated the idea of predicting RMSD from atomic coordinates and built the initial neural network. R.J.L.T., A.M.W., S.E., and M.K. performed and analyzed ARES experiments. A.M.W. generated candidate structural models, with assistance from R.R. R.J.L.T., A.M.W., R.D., and R.O.D. interpreted results. R.J.L.T., A.M.W., and R.O.D. wrote the paper, with input from all authors. **Competing interests:** Stanford University has filed a provisional patent application related to this work. R.J.L.T. is the founder of Atomic AI, an artificial intelligence–driven rational design company. R.D. has received honoraria for seminars at Ribometrix and Pfizer. **Data and materials availability:** Code for the ARES network and data analysis is available on Zenodo (*32*–*34*). All generated models, submitted blind predictions, and other ARES predictions that support the findings of this study are also available at the Stanford Digital Repository (*35*, *36*). The trained ARES model is available at http://drorlab.stanford.edu/ares.html as a web server.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6558/1047/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S5
References (*37*–*65*)
MDAR Reproducibility Checklist

31 August 2020; accepted 14 July 2021
10.1126/science.abe5650

Caltech Division of Chemistry & Chemical Engineering

Assistant Professor of Chemistry or Biochemistry
Division of Chemistry and Chemical Engineering
California Institute of Technology
https://applications.caltech.edu/job/chemistry

The Division of Chemistry and Chemical Engineering at the California Institute of Technology invites applications for a **tenure-track faculty position at the assistant professor level in all areas of chemistry and biochemistry.** Candidates with strong commitments to research and teaching excellence are encouraged to apply. The term of the initial untenured appointment is four years, and the appointment is contingent upon completion of all requirements for a Ph.D. in chemistry, biochemistry, or a related field. Exceptionally well-qualified applicants at the tenured level may also be considered.

Interested candidates should apply at **https://applications.caltech.edu/job/chemistry.** Applications should include a curriculum vita, publication list, teaching statement, a description of proposed research, and three letters of recommendation. Applicants should submit a diversity and inclusion statement that discusses past and/or anticipated contributions to improving diversity, equity, and inclusion in the areas of research, teaching, and/or outreach. Applications should be received by **October 1, 2021**.

We are an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to age, race, color, religion, sex, sexual orientation, gender identity, or national origin, disability status, protected veteran status, or any other characteristic protected by law.

Caltech Division of Chemistry & Chemical Engineering

Assistant Professor of Chemical Engineering
Division of Chemistry and Chemical Engineering
California Institute of Technology
https://applications.caltech.edu/job/chemeng

The Division of Chemistry and Chemical Engineering at the California Institute of Technology invites applications for a **tenure-track faculty position at the assistant professor level in the area of chemical engineering.** Candidates with strong commitments to research and teaching excellence are encouraged to apply. The term of the initial appointment is four years and is contingent upon completion of all requirements for a Ph.D. in chemical engineering or in a related field.

Interested candidates should apply at **https://applications.caltech.edu/job/chemeng.** Applications should include a curriculum vita, publication list, teaching statement, a description of proposed research, and three letters of recommendation. Applicants should submit a diversity and inclusion statement that discusses past and/or anticipated contributions to improving diversity, equity, and inclusion in the areas of research, teaching, and/or outreach. Applications should be received by **November 1, 2021**.

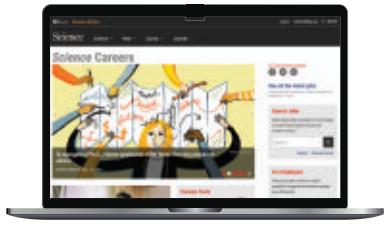
We are an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to age, race, color, religion, sex, sexual orientation, gender identity, or national origin, disability status, protected veteran status, or any other characteristic protected by law.



Science Careers helps you advance your career. Learn how!

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.
- Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
- Complete an interactive, personalized career plan at "my IDP."
- Visit our Employer Profiles to learn more about prospective employers.
- Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free



SCIENCECAREERS.ORG

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SMITHSONIAN INSTITUTION FELLOWSHIP PROGRAM

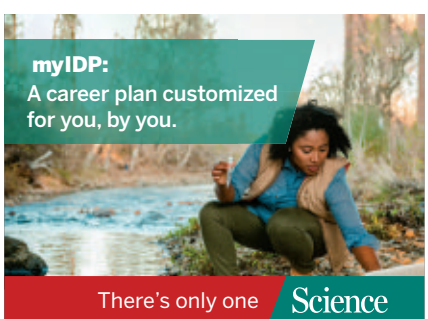
GRADUATE STUDENT, PREDOCTORAL, POSTDOCTORAL, AND SENIOR FELLOWSHIPS in Animal behavior, ecology, and environmental science; including an emphasis on the tropics; Earth sciences and paleobiology; Evolutionary and systematic biology; History of science and technology; Tenable in residence at the Smithsonian facilities. Stipends and tenure vary.

Deadline: **November 1** annually.

Contact: **Office of Fellowships**, Smithsonian Institution, Washington, D.C.

E-mail: **FellowsSI@si.edu**

https://www.si.edu/ofi



Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.

myIDP

Visit the website and start planning today!

myIDP.sciencecareers.org

Science Careers AAAS In partnership with:

UCSF University of California San Francisco

UIC

BURROUGHS WELLCOME FUND

University of Massachusetts Medical School UMASS

FASEB Federation of American Societies for Experimental Biology

utmb Health

Working together to work wonders

Biocontainment Research Associate II or III

A Biocontainment Research Associate II or III Position is available immediately in the laboratory of Dr. Thomas Geisbert to support studies evaluating vaccines and treatments against BSL-3 virus including SARS-CoV-2 and BSL-4 viruses (e.g., Ebola, Lassa, Nipah) and to support work characterizing the pathogenesis of these agents in animal models. Applicants must have a Bachelor's degree in basic science or equivalent or Master's degree in basic science or equivalent. The applicant must meet all State and Federal requirements necessary for working with Select Agents.

The Biocontainment Research Associate II applicant must have at least two years of significant experience in performing cell culture, virology, and immunology procedures.

The Biocontainment Research Associate III applicant must have at least two years of significant experience in performing procedures on laboratory animals.

Duties

• Performs, coordinates, and conducts standard research experiments following technical instructions in BSL-2, BSL-3, and BSL-4 work environments.

• Performs functions of research projects utilizing hazardous viruses that require BSL-3 and BSL-4 containment.

• The Biocontainment Research Associate II Performs technical services including but not limited to maintaining cell cultures, isolating peripheral blood mononuclear cells, plaque assays, virus neutralization assays, immunoblots, ELISA, RNA extractions, PCR assays, multiplex assays, flow cytometry, tissue sampling and processing, and clinical pathology assays (hematology, clinical chemistry, blood coagulation). Also, assists with necropsies and processes specimens for histology and immunohistochemistry.

• The Biocontainment Research Associate III Performs technical services including but not limited performing a variety of procedures on laboratory animals including exposure to viruses, blood collection, tissue collection, vaccination, treatment, euthanasia, and necropsy. Other technical services include but are not limited to performing clinical pathology assays (hematology, clinical pathology, blood coagulation) and processing blood and tissue samples from infected animals.

• Contributes to scientific analysis of projects.

• Writes and reviews methods sections of manuscripts.

• Orders and maintains lab inventory of supplies, animals, and equipment.

• Maintains records of all tests performed and data collected, compiles and tabulates data, and provides analysis of results.

• Maintains knowledge of BSL-2, BSL-3, and BSL-4 operations.

• Maintains independent access to BSL-3 and BSL-4 operations.

• Adheres to internal controls established for the department.

• Performs related duties as required.

Preferred Qualifications

The Biocontainment Research Associate II- More than two years of experience performing in performing cell culture, virology, and immunology procedures is preferred. More than two years of experience working with infectious agents in BSL-3 or BSL-4 is preferred. The ideal candidate would have a DOJ clearance and more than two years of BSL-3 or BSL-4 experience. The applicant must have at least two years of significant experience in performing cell culture, virology, and immunology procedures.

The Biocontainment Research Associate III -More than two years of experience performing procedures including phlebotomy on laboratory animals is preferred. LAT or LATG certification is preferred. More than two years of experience working with infectious agents in BSL-3 or BSL-4 is preferred. The ideal candidate would have a DOJ clearance, LAT or LATG certification, and more than two years of BSL-3 or BSL-4 experience working with laboratory animals.

Supervisor

Thomas Geisbert, Ph.D.

Salary

Salary range: $37,920 -$84,125 dependent on experiences

Location of Position

UTMB, Main Campus (Galveston National Laboratory)

Equal Opportunity Statement

"UTMB Health strives to provide equal opportunity employment without regard to race, color, national origin, sex, age, religion, disability, sexual orientation, gender identity or expression, genetic information, veteran status, or any other basis protected by institutional policy or by federal, state or local laws unless such distinction is required by law. As a Federal Contractor, UTMB Health takes affirmative action to hire and advance women, minorities, protected veterans and individuals with disabilities."

UTMB has several highly interactive research centers, biomedical institutes, and a national biocontainment laboratory with excellent infrastructure to conduct research at BSL2, -3 and -4 on diverse animal models of infectious diseases. The Department, with 31 full-time faculty members, is ranked among the top of its peer departments in NIH funding.

Interested candidates should apply to Requisition # 2103937, Biocontainment Research Associate II or Requisition # 2103526, Biocontainment Research Associate III via the UTMB careers website at https://www.utmb.edu/hr/careers/

ScienceCareers online @sciencecareers.org

By **Ashley M. Bourke**

Waking up to my sleep disorder

Moments before I presented my first seminar as a postdoc to my lab group, my eyelids grew heavy and my sense of alertness dissolved into a dreamlike state. It was a sleep attack—a result of narcolepsy, which I was diagnosed with 2 years earlier. I apologized for my incessant yawning and continued with my talk despite an overwhelming urge to sleep. It's one of many moments when my narcolepsy made my career difficult. But ever since I learned of a renowned scientist who has narcolepsy, I've known I can succeed in science, too.

I've fought the urge to sleep for much of my life. At a young age, I fell asleep in bizarre places—on the front lawn, in a pile of stuffed animals—and in school I was frequently reprimanded for yawning. Fearing others would think I was lazy, I got in the habit of working harder on my schoolwork than everyone else. As an undergraduate, I relied on energy drinks to keep me awake through long hours studying. But when exams were over, I'd stay in bed for days to recuperate.

Graduate school was more difficult because my work was never "done." There was always a new paper to read or an experiment to plan. Shortly after passing my Ph.D. qualifying exam, I worked in the lab every day for more than 30 days straight. That sapped what little energy I had in the first place.



"I ... feared my career in science had reached a dead end."

The stress and exhaustion exacerbated my sleep issues. On more than one occasion, I had to postpone planned experiments and rush home to take a nap. During conversations with my mentor, my stream of thought would sometimes run dry midsentence and I would have to ask him what I was talking about.

I thought depression caused my brain fog and fatigue, so I started seeing a psychiatrist. After 2 years, and multiple medications, I was no better. So she had me complete a sleep questionnaire, which led to a referral to a sleep clinic and a series of sleep tests.

Shortly thereafter I received my diagnosis: narcolepsy. The news was a relief because it meant I might finally receive appropriate treatment and learn how to manage my symptoms. But I also feared my career in science had reached a dead end. As a neuroscientist, how could I uncover the inner workings of the brain with one that is "broken"? How could I succeed if I had to put sleep hygiene ahead of science?

Then I stumbled on an article about George Church, a prominent Harvard University geneticist who has narcolepsy. Calling his condition "a feature, not a bug," he said his daytime bouts of sleep are opportunities for creative, quasi-conscious thought, when he's hatched some of his best ideas. Reading about his condition was empowering. It gave me a sense of belonging, one that went beyond "sure, you can enter our circle" to "you're in our circle and we're better for it."

I wrote Dr. Church an email, telling him how much his perspective meant to me. To my surprise, he sent a detailed reply. "Lecture classes were very hard for me," he wrote. "I even flunked out of my first attempt at Ph.D." His words helped me cultivate a fierce sense of self-compassion. I began to accept that my best can look different from day to day and that I might struggle with some aspects of academia more than others.

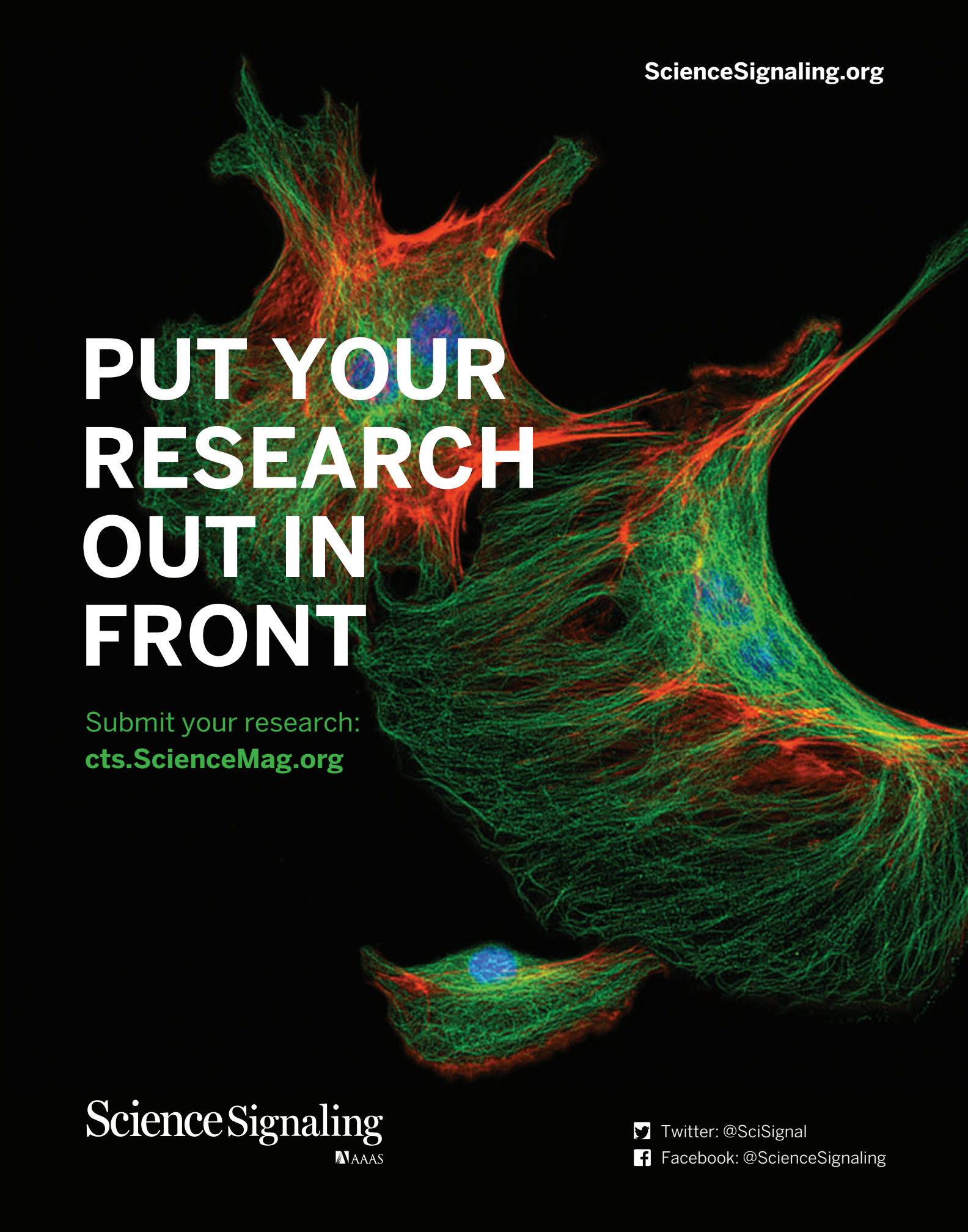
I also implemented many of the strategies he recommended for dealing with narcolepsy, including waiting until evening to eat a large meal and breaking up the day physically with a mixture of standing, pacing, and sitting. I registered with my university's disability office and made requests for workplace accommodations. The university bought me a standing desk, which helped me feel more alert, and I received permission to take naps at work. I made similar requests when I started a postdoc last year.

I still harbor fears of seeming lazy, and I worry that the brain fog I suffer from will one day settle and never lift. But these fears do not stop me from continuing to pursue my goal of managing my own lab one day. Thanks to Dr. Church, I know it's possible—as long as I take care of myself. ■

Ashley M. Bourke is a postdoc at the Max Planck Institute for Brain Research. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

ScienceSignaling.org



PUT YOUR RESEARCH OUT IN FRONT

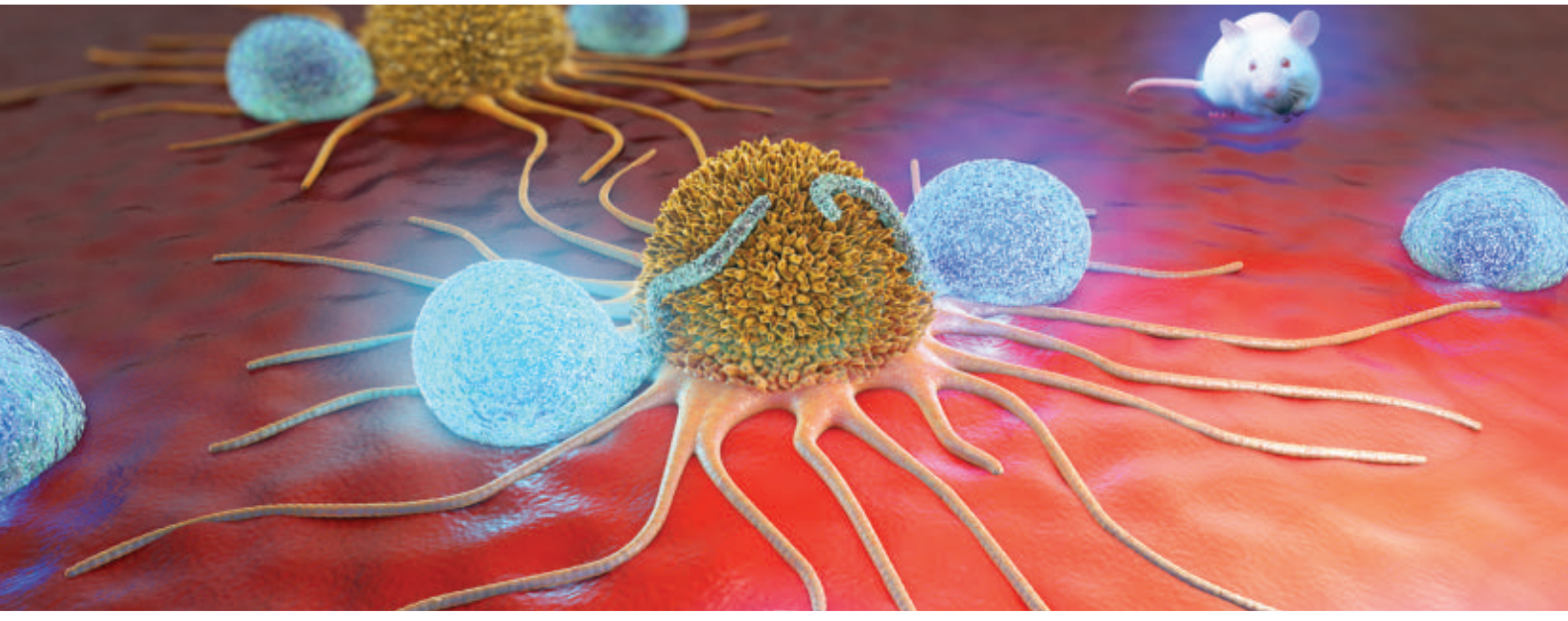
Submit your research:
cts.ScienceMag.org

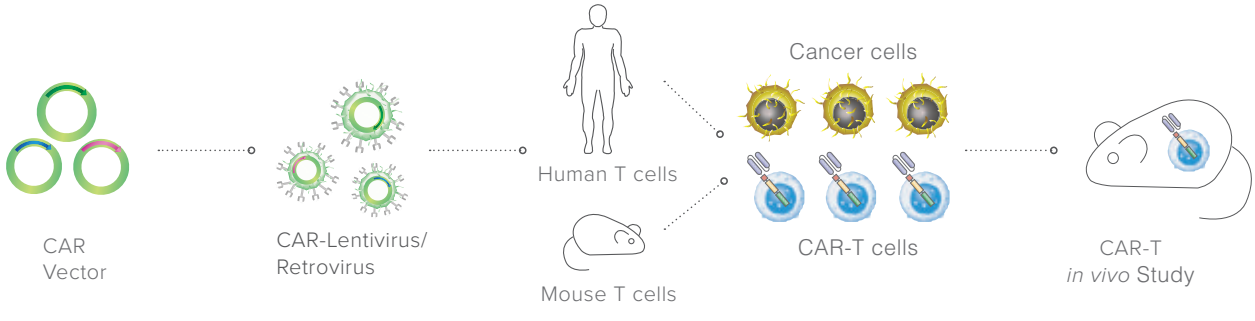
Science Signaling
AAAS

Twitter: @SciSignal
Facebook: @ScienceSignaling

CAR-T/CAR-NK Animal Studies

Testing *In Vivo* Efficacy & More for Cell Therapy Research





Animal Research Custom Services

Custom services available.

- IVIS imaging
- Toxicology
- PK/PD models
- Immunohistochemistry (IHC)
- Cytokine panels
- Tumor volume and weight
- Histopathology
- Genomic parameters
- Cell phenotyping

ProMab Biotechnologies offers a fully comprehensive in vivo functionality service to analyze your cell therapy product using xenograft NSG mice models. CAR-T cells can be made to expedite preclinical research studies in xenogeneic or syngeneic mouse models. ProMab has developed patient-derived xenograft models of different types of cancers to study and validate the efficacy of cell therapy products against solid and hematological tumors and its microenvironment.

All products are for research only

Discover more | www.promab.com

ProMab Biotechnologies, Inc.
2600 Hilltop Dr, Building B, Suite C320, Richmond, CA 94806
1.866.339.0871 | info@promab.com
SCAN ME