How the world's forests shape its weather p. 1302

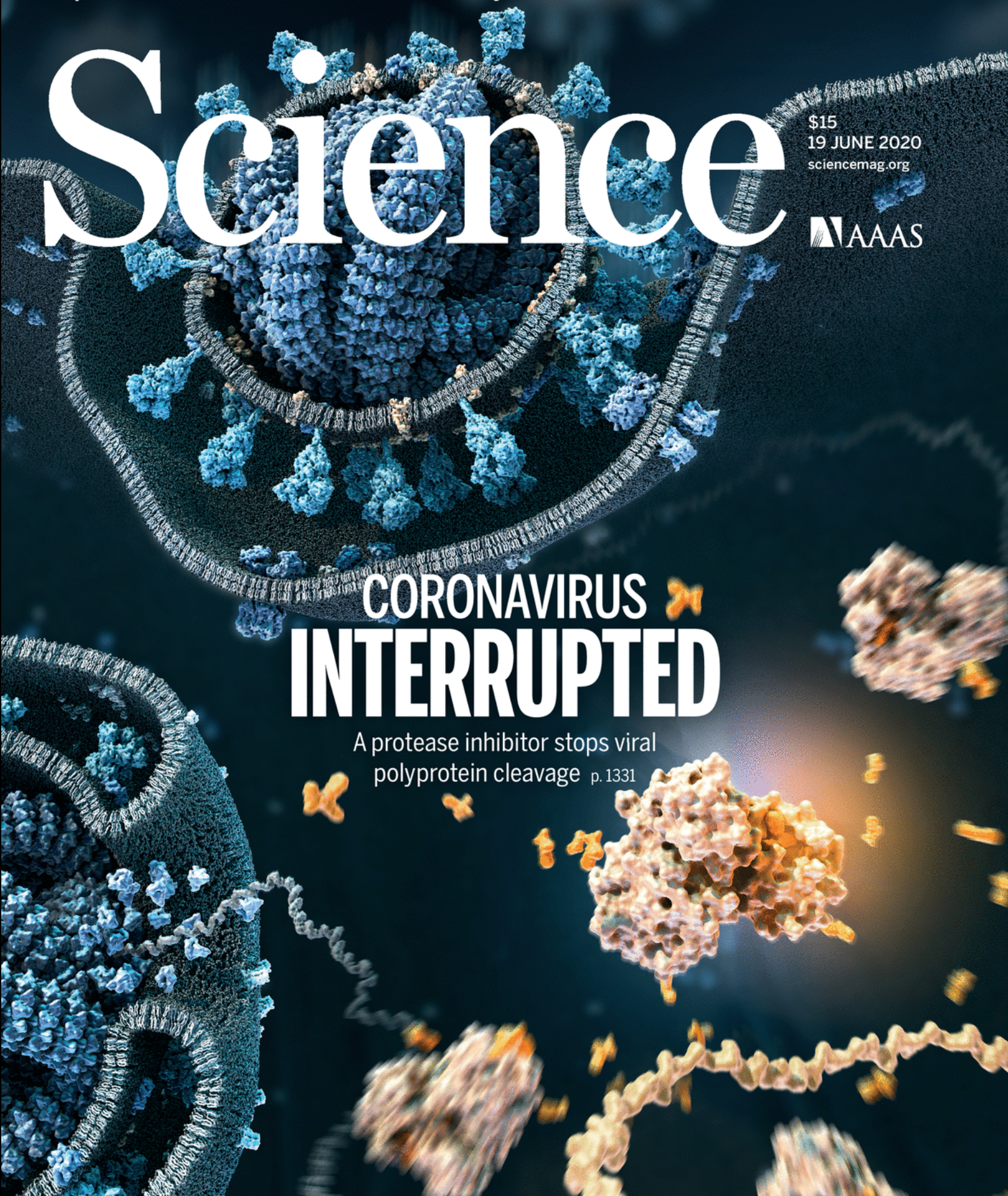
Assessing perovskite solar cell stability pp. 1309 & 1328

Biodiversity change after forest loss p. 1341

Science

$15
19 JUNE 2020
sciencemag.org

AAAS



CORONAVIRUS INTERRUPTED

A protease inhibitor stops viral polyprotein cleavage p. 1331

READY FOR A TRULY LIFE-CHANGING MOMENT? APPLY NOW.

The *Science* & SciLifeLab Prize for Young Scientists is an annual prize awarded to early-career scientists. The prize is presented in four categories: Cell and Molecular Biology; Genomics, Proteomics, and Systems Biology Approaches; Ecology and Environment; and Molecular Medicine.

As a winner, you will have your essay published by *Science*, win up to USD 30,000 and be invited to Sweden where you will receive your award, present your research and meet with leading scientists in your field.

Get ready for a life-changing moment in your scientific career – apply now, deadline is July 15!

SCIENCEPRIZE.SCILIFELAB.SE

Knut och Alice Wallenbergs Stiftelse

Science AAAS

SciLifeLab

2019 winners (from left): Longzhi Tan, Zibo Chen, Barbara Klump and Humsa Venkatesh.

CONTENTS

19 JUNE 2020 • VOLUME 368 • ISSUE 6497

1302

NEWS

IN BRIEF

1292 News at a glance

IN DEPTH

1295 Pandemic vaccines are about to face the real test
U.S. and global efforts are taking different approaches to key efficacy trials of candidates *By J. Cohen*

1296 Can phone apps slow the spread of the coronavirus?
Digital contact tracing is growing—and facing its first real-world tests *By K. Servick*

1298 U.S. academic research funding stays healthy despite pandemic
Labs reopening even as campuses struggle
By J. Mervis
PODCAST

1299 Incest in ancient Ireland suggests an elite ruled early farmers
DNA from massive Newgrange tomb reveals a practice linked to royalty around the world
By A. Curry

1300 New tensions dim hopes for salvaging Iran nuclear deal
Monitoring agency seeks clarity about clandestine R&D *By R. Stone*

1301 NIH requires disclosure about sexual harassment by grantees
Changes aimed in part at making sure institutions don't "pass the harasser" *By J. Kaiser*

FEATURES

1302 Weather makers
Forests supply the world with rain. A controversial Russian theory claims they also make wind *By F. Pearce*

INSIGHTS

PERSPECTIVES

1306 The origins of flowering plants and pollinators
New research raises questions about when flowering plants and their pollinators evolved
By C. J. van der Kooi and J. Ollerton

1309 Perovskite solar cells take a step forward
A new encapsulation technique helps move a photovoltaic toward commercialization
By E. J. Juarez-Perez and M. Haro
RESEARCH ARTICLE p. 1328

1310 Dating the emergence of human pathogens
Ancient genomes can narrow the search for the sources of zoonotic transmissions
By S. Y. W. Ho and S. Duchêne
REPORT p. 1367

1311 Getting a grip on touch receptors
Meissner corpuscles are anatomically complex mechanosensors critical for tactile acuity
By K. Marshall and A. Patapoutian
RESEARCH ARTICLE p. 1330

1312 Closing the radical gap in chemical synthesis
Unstable radical intermediates are harnessed in a microfluidic electrochemical cell *By J.-Q. Liu* et al.
REPORT p. 1352

1314 Drug modulation by nuclear condensates
Concentration of antineoplastic agents into spatial compartments influences activity *By A. D. Viny and R. L. Levine*
REPORT p. 1386

1315 Using information theory to decode network coevolution
Communication clashes shape the coevolution of insect-plant ecosystems *By R. Solé*
REPORT p. 1377

POLICY FORUM

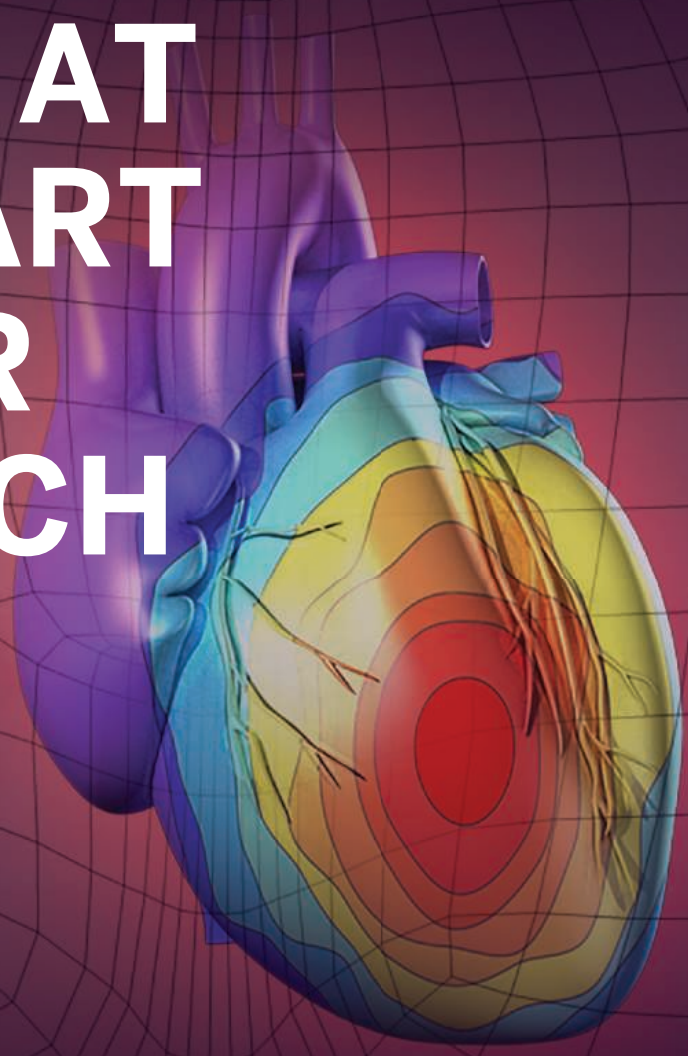
1317 Understanding persistent gender gaps in STEM
Does achievement matter differently for men and women?
By Joseph R. Cimpian et al.

PHOTO: BRUSINI AURELIEN/HEMIS/ALAMY STOCK PHOTO

Published by AAAS

ScienceTranslationalMedicine.org

PUT HUMAN HEALTH AT THE HEART OF YOUR RESEARCH



Submit your research:
cts.ScienceMag.org

Science Translational Medicine
AAAS

Twitter: @ScienceTM
Facebook: @ScienceTranslationalMedicine

BOOKS *ET AL.*

1320 Creating a culture of change
Urging action, a new film paints a harrowing portrait of female scientists' experiences in academia *By D. Riley*

1321 What to do with our days
When wielded correctly, boredom can be a powerful tool *By E. Westgate*

LETTERS

1322 Academic societies' role in curbing police brutality
By P. Phillips and M. B. Weissman

1322 Pandemics' historical role in creating inequality
By L. M. Dávalos et al.

1323 Recent immigrants at increased pandemic risk
By P. Galanaud and A. Galanaud

RESEARCH

IN BRIEF

1324 From *Science* and other journals

REVIEW

1327 Forest and climate
Climate-driven risks to the climate mitigation potential of forests *W. R. L. Anderegg* et al.
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAZ7005

RESEARCH ARTICLES

1328 Solar cells
Gas chromatography–mass spectrometry analyses of encapsulated stable perovskite solar cells *L. Shi* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABA2412
PERSPECTIVE p. 1309

1329 Neuroscience
Manipulating synthetic optogenetic odors reveals the coding logic of olfactory perception *E. Chong* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABA2357

1330 Neuroscience
Meissner corpuscles and their spatially intermingled afferents underlie gentle touch perception *N. L. Neubarth* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABB2751
PERSPECTIVE p. 1311

PHOTO: MICHAEL DANTAS/AFP/GETTY IMAGES

1331 Coronavirus
Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease *W. Dai* et al.

1335 Influenza
Different genetic barriers for resistance to HA stem antibodies in influenza H3 and H1 viruses *N. C. Wu* et al.

1341 Biodiversity change
Landscape-scale forest loss as a catalyst of population and biodiversity change *G. N. Daskalova* et al.

REPORTS

1347 Metallurgy
Making ultrastrong steel tough by grain-boundary delamination *L. Liu* et al.

1352 Electrochemistry
Microfluidic electrochemistry for single-electron transfer redox-neutral reactions *Y. Mo* et al.
PERSPECTIVE p. 1312

1357 Geophysics
3D fault architecture controls the dynamism of earthquake swarms *Z. E. Ross* et al.

1362 Coronavirus
Rapid implementation of mobile technology for real-time epidemiology of COVID-19 *D. A. Drew* et al.

1367 Measles
Measles virus and rinderpest virus divergence dated to the sixth century BCE *A. Düx* et al.
PERSPECTIVE p. 1310

1371 Immunometabolism
T cells with dysfunctional mitochondria induce multimorbidity and premature senescence *G. Desdín-Micó* et al.

1377 Chemical ecology
Information arms race explains plant-herbivore chemical communication in ecological communities *P. Zu* et al.
PERSPECTIVE p. 1315

1381 Metalloenzymes
Structural evidence for a dynamic metallocofactor during N_2 reduction by Mo-nitrogenase *W. Kang* et al.

1386 Phase separation
Partitioning of cancer therapeutics in nuclear condensates *I. A. Klein* et al.
PERSPECTIVE p. 1314

DEPARTMENTS

1290 Editorial
COVID-19 and cancer *By Norman E. Sharpless*

1291 Editorial
Combating sexual harassment *By Carrie D. Wolinetz, Michael S. Lauer, Francis S. Collins*

1394 Working Life
Undergrads in charge *By Akira Nishii*

ON THE COVER

Upon infection, SARS-CoV-2 (blue) uses host machinery to produce polyproteins for replication. The viral main protease (yellow) plays a key role in polyprotein cleavage and thus is a potential drug target. Researchers have designed two antiviral compounds that inhibit activity of the main protease to prevent viral replication. Preclinical studies show that these compounds are promising therapeutic candidates. See page 1331. *Illustration: C. Bickel/*Science; *Data: PDB ID 6M0K (SARS-CoV-2 main protease)*

Science Careers 1393

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 © 2020 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.

Published by AAAS

EDITORIAL

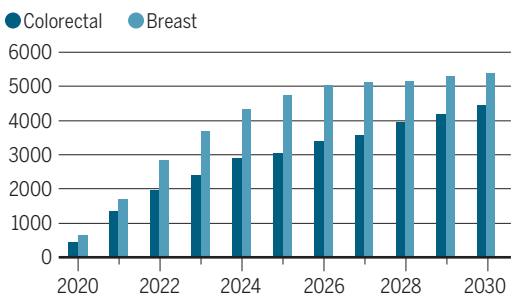
COVID-19 and cancer

Norman E. Sharpless is director of the U.S. National Cancer Institute, Bethesda, MD, USA. norman.sharpless@nih.gov

With the spread of coronavirus disease 2019 (COVID-19), countries and states have instituted lockdowns. These decisions have been difficult and are sometimes described as benefiting the public health at the expense of the economy. Fear of contracting the coronavirus in health care settings has dissuaded people from screening, diagnosis, and treatment for non–COVID-19 diseases. The consequences for cancer outcomes, for example, could be substantial. What can be done to minimize this effect?

Cancer is a complex set of diseases whose prognoses are influenced by the timing of diagnosis and intervention. In general, the earlier one receives cancer treatment, the better the results. There already has been a steep drop in cancer diagnoses in the United States since the start of the pandemic, but there is no reason to believe the actual incidence of cancer has dropped. Cancers being missed now will still come to light eventually, but at a later stage ("upstaging") and with worse prognoses. At many hospitals, so-called "elective" cancer treatments and surgeries have been deprioritized to preserve clinical capacity for COVID-19 patients. For example, some patients are receiving less intense chemotherapy and/or radiotherapy, and in other cases, patients' operations to remove a newly detected tumor are being delayed. There can be no doubt that the COVID-19 pandemic is causing delayed diagnosis and suboptimal care for people with cancer.

Modeled cumulative excess deaths from colorectal and breast cancers, 2020 to 2030*



What will be the likely impact of the pandemic on cancer mortality in the United States? Modeling the effect of COVID-19 on cancer screening and treatment for breast and colorectal cancer (which together account for about one-sixth of all cancer deaths) over the next decade suggests almost 10,000 excess deaths from breast and colorectal cancer deaths; that is, a ~1% increase in deaths from these tumor types during a period when we would expect to see almost 1,000,000 deaths from these two diseases types.* The number of excess deaths per year would peak in the next year or two. This analysis is conservative, as it does not consider other cancer types, it does not account for the additional nonlethal morbidity from upstaging, and it assumes a moderate disruption in care that completely resolves after 6 months. It also does not account for regional variations in the response to the pandemic, and these effects may be less severe in parts of the country with shorter or less severe lockdowns.

Beyond clinical care, the COVID-19 pandemic has caused an unprecedented disruption throughout the cancer research community, shuttering many labs and slowing down cancer clinical trial operations. Many scientists and clinicians are pivoting their cancer research activities to study the impact of SARS-CoV-2 on cancer. The scientific community must ensure that this pause is only temporary, because trials are the only way to make progress in developing new therapies for cancer. Given the long timeline between basic cancer research and changes to cancer care, the effects of pausing research today may lead to slowdowns in cancer progress for many years to come.

Collective action by the clinical and research communities and by governmental agencies can mitigate this potentially substantial impact. The U.S. National Cancer Institute (NCI), for example, has started to address this challenge (see www.cancer.gov). The NCI has worked with the U.S. Food and Drug Administration to increase flexibility and support for clinical trials during the pandemic. For example, allowances have been made to accept "remote" informed consent, and other protocol deviations. In addition, the NCI has announced several new clinical trials and funding opportunities aimed at addressing the relationship between COVID-19 and cancer. Of particular note is the NCI COVID-19 in Cancer Patients Study, a prospective longitudinal study that will collect blood samples, imaging, and other data to understand how COVID-19 affects cancer patients.

Clearly, postponing procedures and deferring care as a result of the pandemic was prudent at one time, but the spread, duration, and future peaks of COVID-19 remain unclear. However, ignoring life-threatening non–COVID-19 conditions such as cancer for too long may turn one public health crisis into many others. Let's avoid that outcome.

–Norman E. Sharpless

*See supplementary materials (science.sciencemag.org/content/368/6497/1290/suppl/DC1).

10.1126/science.abd3377

CREDITS: (PHOTO) NATIONAL CANCER INSTITUTE; (GRAPH) V. ALTOUNIAN/SCIENCE; (DATA) NATIONAL CANCER INSTITUTE

Published by AAAS

EDITORIAL

Combating sexual harassment

Sexual harassment, including gender harassment, presents an unacceptable barrier that prevents women from achieving their rightful place in science, and robs society and the scientific enterprise of diverse and critical talent. As the largest single funder of biomedical research in the world, the U.S. National Institutes of Health (NIH) bears a responsibility to take action to put an end to this behavior. In 2019, the NIH began to bolster its policies and practices to address and prevent sexual harassment. This included new communication channels to inform the agency of instances of sexual harassment related to NIH-funded research. This week, the NIH announces a change that will hold grantee institutions and investigators accountable for this misconduct, to further foster a culture whereby sexual harassment and other inappropriate behaviors are not tolerated in the research and training environment.

Last year, an Advisory Committee to the Director (ACD) of the NIH presented a report and recommendations to end sexual harassment. A major theme of this report was the need for increased transparency and accountability in the reporting of professional misconduct, especially sexual harassment. The cases of sexual harassment that surfaced in the wake of the U.S. National Academies of Sciences, Engineering, and Medicine (NASEM) 2018 report highlighted a substantial gap in the NIH's oversight of the research enterprise: There was no straightforward mechanism for the agency to learn of sexual harassment or other misconduct taking place at grantee institutions in the context of NIH-funded research. It was not uncommon for the NIH to discover such cases through the media, amid rightful public outcry. Holding institutions and investigators accountable for this behavior was challenging.

Over the past year, the NIH has established communication channels for anyone—institutions, targets of harassment, bystanders—to inform the agency about instances when sexual harassment compromised the conduct of an NIH-funded project at a grantee institution. Although voluntary provision of information is important and has improved awareness of, and increased NIH actions on, sexual harassment cases, it does not fully cover issues related to the required reporting of a change in principal investigator (PI) or the transfer of a grant by an NIH grantee institution. Two important areas required more attention: differentiating between instances where an institution removed a PI from a grant because of findings or concerns of sexual harassment from other reasons (such as medical leave or job change); and preventing "passing the harasser," in which a scientist who changed institutions could evade the consequences of findings of sexual harassment.

To close these gaps, the NIH has issued new guidance to grantees that sets clear expectations for reporting to the NIH when a grantee institution has a finding of sexual harassment of a PI named on an NIH grant. The notice also makes it clear that the NIH expects its grant recipients who request changes in either investigators or movement of a grant to a new recipient institution to promptly inform the agency, whether changes are related to concerns about safety and/or work environments (e.g., because of concerns about harassment, bullying, retaliation, or hostile working conditions) (see https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-124.html). This includes situations where a PI, or other senior personnel, are removed from a grant during the investigation of a serious allegation. The NIH will use such information in making decisions related to grant funding, including putting into place appropriate institutional oversight mechanisms or remedies, or deciding whether to grant institutional requests about personnel on grant awards, thus connecting information to direct consequences. This guidance should also unmask NIH-funded harassers who seek to change institutions without revealing their track record, whether the investigator seeks to take an existing grant to a new institution or leaves an existing grant behind for a new PI to take it over. These questions would have to be answered and would reveal any past misconduct to the new institution.

> "...National Institutes of Health (NIH) bears a responsibility to... put an end to this behavior."

The NASEM report noted that risk factors for climates and cultures that fomented sexual harassment included both the perception that reports of sexual harassment would be ignored and the perception that there would be no consequence for investigators. It is our hope that this new guidance sends a clear message to dispel those perceptions.

–Carrie D. Wolinetz, Michael S. Lauer, Francis S. Collins

Carrie D. Wolinetz is the associate director for Science Policy and acting chief-of-staff to the NIH Director, National Institutes of Health, Bethesda, MD, USA. carrie.wolinetz@nih.gov

Michael S. Lauer is the deputy director for Extramural Research, National Institutes of Health, Bethesda, MD, USA. michael.lauer@nih.gov

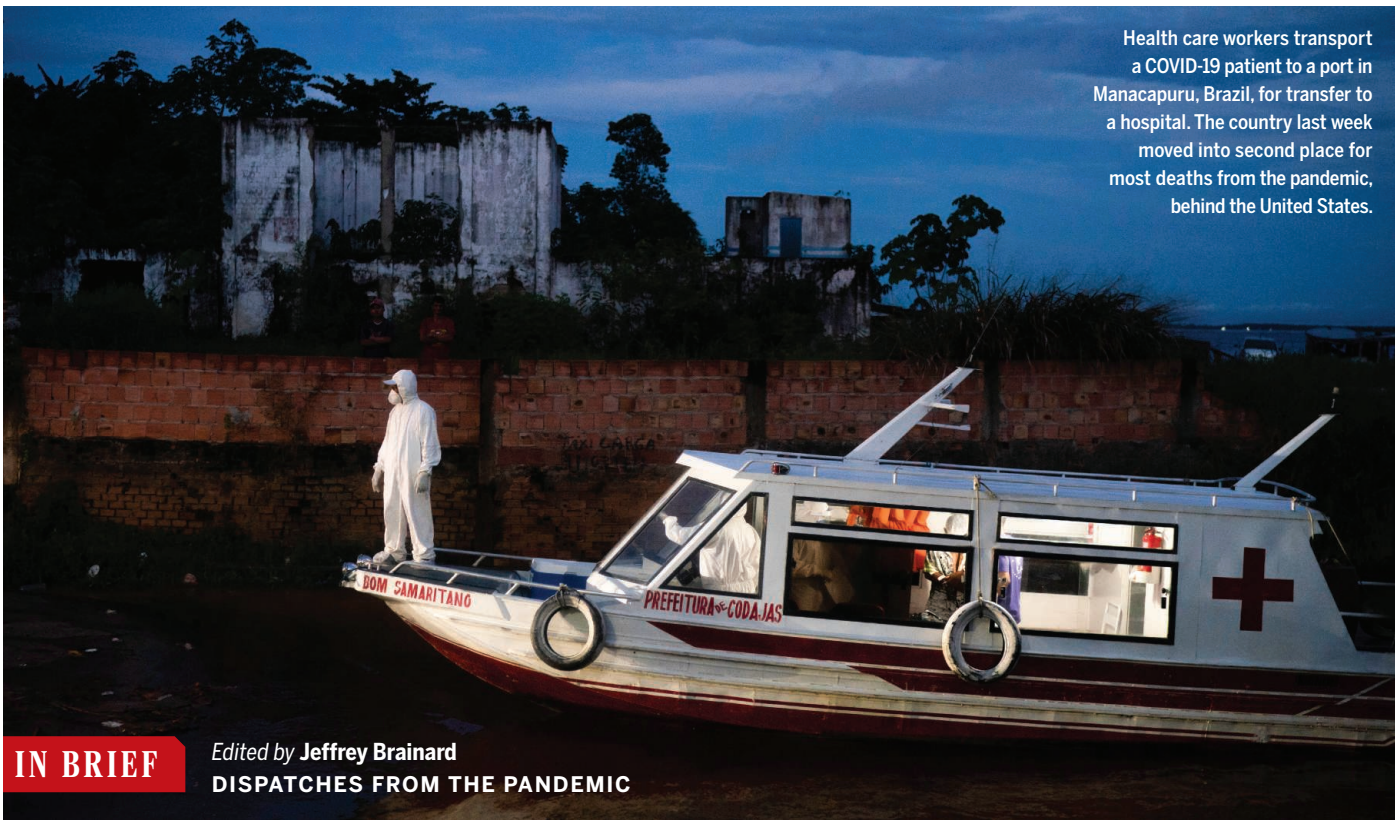
Francis S. Collins is the director of the National Institutes of Health, Bethesda, MD, USA. collinsf@od.nih.gov

Published online 11 June 2020; 10.1126/science.abd2644

Published by AAAS

NEWS

> “I don’t want more diversity and inclusion seminars. [They] are used to provide a shield to institutions.”
>
> **Astrophysicist Brian Nord** of Fermilab, an organizer of the global #ShutDownSTEM protests on 10 June, calling for effective plans to address racism in science.



Health care workers transport a COVID-19 patient to a port in Manacapuru, Brazil, for transfer to a hospital. The country last week moved into second place for most deaths from the pandemic, behind the United States.

PHOTO: FELIPE DANA/AP PHOTO

IN BRIEF

Edited by **Jeffrey Brainard**

DISPATCHES FROM THE PANDEMIC

FDA backs off malaria drugs

REGULATION | The U.S. Food and Drug Administration (FDA) on 15 June revoked its emergency use authorization for hydroxychloroquine and chloroquine to treat COVID-19. The two antimalaria drugs, touted by President Donald Trump and others as potential game-changers, have failed in recent rigorous clinical trials to prevent disease in newly infected people or help those with symptoms. Former FDA leaders had criticized the agency’s decision to authorize distribution of the drugs from a national stockpile for use on COVID-19 patients, asserting it was based on political pressure, not scientific evidence. The agency said that a review showed the two drugs were “unlikely to be effective in treating COVID-19” and that “serious cardiac adverse events and other serious side effects” outweighed the drugs’ known and potential benefits. FDA’s action still allows the drugs to be used in clinical trials of COVID-19 patients.

Steroid treatment shows promise

BIOMEDICINE | Researchers in the United Kingdom announced this week they have found a cheap, safe drug that could save the lives of patients severely sick with COVID-19: the corticosteroid dexamethasone. The U.K. Recovery trial, one of the biggest efforts to find drugs to treat the disease, included 2104 patients treated with 6 milligrams of dexamethasone for 10 days and compared their outcomes with those of 4321 patients receiving usual care. They found the treatment reduced deaths by one-third in ventilated patients and by one-fifth in patients receiving oxygen only. Scientists announced the results in a press release and said they were working to publish the full results soon.

Antarctic field season postponed

GLACIOLOGY | The U.S. National Science Foundation and the British Antarctic Survey announced last week they would suspend most of their planned Antarctic research during the coming field season to avoid COVID-19 infections. The halt, during Antarctica’s summer from October to March, is especially painful for the International Thwaites Glacier Collaboration, the continent’s largest field project ever. The remote Thwaites is the West Antarctic glacier most at risk of near-term melting. The 2020–21 season was set to be the collaboration’s second on the ice—and perhaps its most critical. So far, both the United States and United Kingdom have prevented infections of the novel coronavirus at their Antarctic stations. Most other nations conducting research on the continent are expected to announce similar steps.

TB, HIV pose some COVID-19 risk

PUBLIC HEALTH | Living with HIV or tuberculosis (TB) increases a person’s likelihood of dying from COVID-19, new data from South Africa suggest—but not

Published by AAAS

dramatically. An analysis presented in a 9 June webinar by Mary-Ann Davies, a public health expert with the Western Cape government, showed HIV infection raises the risk of dying from COVID-19 by 2.75 times, and TB by 2.58. Those increases are smaller than those posed by other risk factors: The death risk for South Africans in their 50s was 10 times higher than for those younger than 40, for instance, and diabetics were between four and 13 times more likely to die than nondiabetics. South Africa has the world's highest rates of infection with HIV, which undermines the immune system, and TB, which can damage the lungs. Some specialists had worried these vulnerabilities would allow the population to be ravaged by COVID-19.

Colombia coca spraying delayed

DRUG CONTROL | The COVID-19 pandemic has at least temporarily derailed a controversial plan by Colombia's government to resume aerial spraying of a chemical used to kill coca crops that feed the global trade in cocaine. A Colombian court ruled in late May that the spraying of the herbicide glyphosate, which some studies have linked to human health and environmental problems, cannot resume until the government consults with affected communities—a process that has been severely disrupted by the ongoing pandemic. Colombia halted its aerial spraying program in 2015 amid health concerns.

Grant cancellation furor grows

OVERSIGHT | Scientists who counsel National Institutes of Health (NIH) Director Francis Collins have joined a chorus of concerns about the 24 April cancellation of a peer-reviewed grant to study bat coronaviruses in China. Last week, at a meeting of NIH's Advisory Committee to the Director, Baylor College of Medicine geneticist Brendan Lee read a statement from 15 of 17 ACD members. It urged Collins to "thoroughly" review the process that led NIH to kill the grant to the EcoHealth Alliance and consider reversing its decision. NIH acted after conservative politicians and commentators suggested, without evidence, that the coronavirus causing the current pandemic escaped from a lab in Wuhan, China, that received funding from the grant and President Donald Trump said he would end it.

SCIENCEMAG.ORG/TAGS/CORONAVIRUS
Read additional *Science* coverage of the pandemic.

FOREIGN INFLUENCE

NIH probe spurs removal of 54 scientists

The U.S. National Institutes of Health (NIH) says 54 scientists at outside institutions have resigned or been fired as a result of its ongoing investigation into grantees who did not disclose their ties to foreign governments. Michael Lauer, head of extramural research, told a top NIH advisory panel last week that Chinese institutions provided the funding for 93% of the 189 scientists his office has investigated since August 2018. That pool is overwhelmingly Asian (82%) and male (82%); Lauer said China targets Asians in its foreign talent recruitment programs. Some 70% had failed to report receipt of a foreign grant, and 54% their participation in a foreign talent program. Despite fears that these hidden relationships are a vehicle for stealing intellectual property, only 9% of cases involved ties to a foreign company and only 4% had an undisclosed foreign patent.

U.S. nanotechnology lags China

POLICY | The United States should overhaul its research priorities in nanotechnology and bolster funding to restore its leadership in the discipline, a panel of the National Academies of Sciences, Engineering, and Medicine said in a quadrennial review published 9 June. After U.S. dominance in the 2000s, China now produces nearly double the annual number of nanotechnology publications and more than three times the annual number of patents. To regain its edge, the United States needs to upgrade facilities, increase funding, and attract more students to the field, the report says. Coordinating federal nanotechnology efforts, the panel said, is challenging in part because it involves 20 agencies and offices, each with its own budget and priorities.

NOAA chief criticized

LEADERSHIP | The acting administrator of the U.S. National Oceanic and Atmospheric Administration (NOAA), Neil Jacobs, violated scientific integrity policies in September 2019 during the Hurricane Dorian response, an independent report commissioned by the agency has concluded. After President Donald Trump falsely claimed that Alabama was threatened by the hurricane—a weather map was altered accordingly in an incident quickly dubbed "Sharpiegate"—local NOAA forecasters correctly stated it was not. Several days later, under pressure from unnamed officials at the Commerce Department, NOAA's parent agency, Jacobs helped write an unsigned statement that rebuked those staff. That politically driven response, the report concludes, violated the agency's guidelines. Jacobs remains NOAA's acting administrator. NOAA has not fired anyone in response to the violation, instead recommending improved training and policies.

PubMed to add preprints

PUBLICATIONS | To accelerate sharing of research findings, managers of PubMed, the database of biomedical abstracts run by the U.S. National Institutes of Health (NIH), have begun an effort to list draft papers that have not undergone peer review. NIH will add papers posted on preprint servers such as bioRxiv if they cover research supported by its funding. The move follows the agency's 2017 decision to encourage investigators to include preprints in proposals and progress reports. The 1-year pilot will first index the flood of preprints on the COVID-19 pandemic. Users seeking only peer-reviewed articles can exclude preprints from search results.

BY THE NUMBERS

12.4

Average hours U.S. high school biology classes devoted to teaching evolutionary processes in 2019, up 25% from 2007. More than 80% of teachers did not agree with teaching intelligent design as an alternative to Darwinism (*Evolution: Education and Outreach*).

Published by AAAS

In this fusion reactor's 3-meter-high vacuum chamber, a plasma can reach temperatures above 10 million degrees Celsius.

ENERGY

Fusion lab gets green light for repairs

The Princeton Plasma Physics Laboratory (PPPL) last week received permission from the U.S. Department of Energy (DOE) to begin a $199 million repair of its main fusion energy experiment, a design known as a tokamak. In 2016, a short circuit in a magnetic coil idled the National Spherical Torus Experiment just 10 weeks after a $94 million upgrade. The problem triggered an investigation and a leadership change at PPPL, DOE's sole fusion lab. In 2022, the rebuilt facility is expected to generate plasmas of deuterium, a heavy isotope of hydrogen, allowing researchers to study conditions in which nuclei can fuse and release energy.

Firms halt facial scanner sales

TECHNOLOGY | After widespread global protests about police killings of black people, three technology giants pledged last week to stop selling facial recognition software. The move comes 2 years after researchers and civil rights organizations criticized the technology, citing studies indicating it misclassified darker skinned faces far more often than white ones, putting people of color at increased risk of racially biased policing. After initially disputing the findings, Amazon on 10 June reversed itself and suspended its software sales for 1 year. IBM announced on 9 June it would abandon the sales, and Microsoft on 11 June suspended selling its software until there is a national law regulating the technology.

Strict laws reduce gun deaths

POLICY RESEARCH | States that restrict children's access to guns, don't allow people to carry concealed weapons without a permit, and don't allow lethal "stand your ground" self-defense can expect to have 11% fewer gun deaths every year than states with more permissive gun laws, according to a study this week in the *Proceedings of the National Academy of Sciences*. The three types of gun laws were common enough from 1980 to 2016 to reveal meaningful statistical relationships.

IN OTHER NEWS

MOON ROVER NASA awarded a $200 million contract to Astrobotic Technology for a lander to carry the first rover it will send to the Moon's surface. The robot, which will scout for ice that could supply water for astronauts, is to arrive at the Moon in 2023.

OPEN-ACCESS DEAL The University of California (UC) system has signed the biggest open-access deal in North America with one of the largest commercial scientific publishers, Springer Nature. The publisher agreed to explore making all articles that UC corresponding authors publish in the *Nature* family of journals free to read on publication starting in 2022.

NOMINEE OPPOSED More than 90 scientists asked a U.S. Senate panel not to confirm President Donald Trump's nominee, toxicologist Nancy Beck, as chair of the Consumer Product Safety Commission. The researchers, including former employees of the Environmental Protection Agency, cite decisions she made as head of EPA's chemical safety office about toxic substances that they say put workers at risk.

PHOTO: ELLE STARKMAN/PPPL COMMUNICATIONS/CC BY

Published by AAAS

IN DEPTH

Mobile teams will help test COVID-19 vaccine candidates, a strategy used to evaluate an experimental Ebola vaccine in the Democratic Republic of the Congo.

PHOTO: DIANA ZEYNEB ALHINDAWI/THE NEW YORK TIMES/REDUX

COVID-19

Pandemic vaccines are about to face the real test

U.S. and global efforts are taking different approaches to key efficacy trials of candidates

By **Jon Cohen**

A Chinese company will turn to Brazil for help. The World Health Organization (WHO) is adopting a strategy forged in a war zone during an Ebola outbreak. And the Trump administration plans to lean on existing infrastructure for testing HIV and flu vaccines. These are the disparate strategies about to be employed in the next and most important stage of the COVID-19 vaccine race: the large-scale, placebo-controlled, human trials needed to prove which of the more than 135 candidates are safe and effective.

Two such efficacy trials plan to start next month, even as the United States and global initiatives struggle to answer major questions, from what it means for a COVID-19 vaccine to work to how to find enough people exposed to the virus so a candidate can be put to a real-world test. Populations that have high levels of viral transmission are a moving target—Wuhan, Seattle, or Milan might once have been a good place to test the mettle of a vaccine, but no longer. And quickly enrolling tens of thousands of properly informed people who meet a trial's entry criteria is a "big lift," says Susan Buchbinder, an epidemiologist at the San Francisco Department of Public Health who runs vaccine trials.

Competition among trial efforts could hinder the global push, says Wayne Koff, who heads the nonprofit Human Vaccines Project and formerly led the HIV vaccine program at the National Institute of Allergy and Infectious Diseases (NIAID). "It's absolutely extraordinary how much has been done in 6 months, but there's an old adage that everybody loves to collaborate unless they want to win." Others, however, don't anticipate conflicts, noting that scientists and officials are sharing information about trial designs and plans. "Each one will contribute differently," says Ana Maria Henao Restrepo, lead representative of WHO's vaccine effort, Solidarity. "I don't see competition."

"Winning," one of U.S. President Donald Trump's favorite terms, is the clear goal of Operation Warp Speed, the U.S. project that aims to start to vaccinate millions of Americans in October and offer shots to 300 million people in the United States by January 2021. After winnowing down vaccine candidates in an opaque process over the past month and committing what could be more than $2 billion to its top choices, Warp Speed plans to enter three to five of them into efficacy trials that will have "harmonized" protocols to streamline oversight and will run analyses in central labs so data can more easily be compared.

The first Warp Speed candidate to launch, from the biotech Moderna, is composed of RNA encoding the coronavirus spike protein from SARS-CoV-2, the virus that causes COVID-19. The candidate's efficacy trial, announced by the company on 11 June, will enroll 30,000 people and take place at established HIV and flu vaccine test sites, primarily at U.S. hospitals and universities, now overseen by Warp Speed. But which of those brick-and-mortar sites will have enough SARS-CoV-2 circulating near them to quickly produce an efficacy signal is uncertain given the shifting distribution of new cases in the United States.

China has an even starker problem: There's currently little transmission to speak of in the country, which has forced Sinovac Biotech, a company based in Beijing, to stage efficacy trials of its vaccine candidate in Brazil, where the COVID-19 epidemic is now raging. With a product com-

Science's COVID-19 coverage is supported by the Pulitzer Center.

Published by AAAS

posed of the entire virus, inactivated with chemicals, Sinovac announced this week it is collaborating with the Butantan Institute, a major research institution in São Paulo that manufactures vaccines. "We are working very hard to start the trial in July," says Sinovac Senior Director Meng Weining.

WHO proposes a different solution for Solidarity's efficacy trials. The agency hasn't yet announced which candidates Solidarity will test, but, unlike Warp Speed—which won't consider Chinese-made vaccines—it is open to products from every country and has made public detailed criteria for how it will prioritize vaccines. To cope with the patchiness of the pandemic, Solidarity will adopt a strategy Henao Restrepo helped develop for Ebola vaccine trials in Guinea in 2015 and, 3 years later, in the Democratic Republic of the Congo (DRC): setting up vaccination teams that can quickly mobilize to localized outbreaks.

"We did this in Congo despite the war," Henao Restrepo says. "It's not the traditional way, and some people think that we are crazy, but we have done it not once but twice." In the DRC, about 20 teams with 15 members each drove around the affected regions and set up temporary sites, vaccinating and following more than 300,000 people.

Warp Speed, which could if needed expand its trials to international sites used for HIV drug and vaccine testing, also plans to form "surge clinics" to quickly recruit people in rural U.S. areas with big outbreaks or pockets of high transmission such as nursing homes. Models driven by machine learning will help Warp Speed forecast where infection will be highest, says Peter Gilbert, a University of Washington, Seattle, biostatistician. "There are risk predictors that account for space and geography and features that are more constant like race, ethnicity, or preexisting conditions," Gilbert says. "It's really complicated."

One of the trickiest issues for trial designers is deciding what, exactly, represents success for a COVID-19 vaccine. "Is it an infection endpoint, a transmission endpoint, preventing moderate disease, or preventing severe disease?" Koff asks.

"There was a lot of debate on that question," for Warp Speed, notes John Mascola, who heads NIAID's Vaccine Research Center and contributes to the project. A COVID-19 vaccine that fails to prevent infection might still provide great benefit if it reduces symptomatic disease, so Warp Speed and Solidarity both ultimately chose that as the primary endpoint of the trials. Trial volunteers who develop fever, headache, dry cough, or other symptoms linked to COVID-19 will be tested for SARS-CoV-2, to see whether more people with confirmed infections develop symptomatic disease in the placebo arm of the trial than among those who received the vaccine.

To detect an efficacy signal, both Warp Speed and Solidarity estimate they will need to give each vaccine to 15,000 to 20,000 people in a population that has a 1% per year incidence of SARS-CoV-2 infection. If the vaccine prevents COVID-19 symptoms at least 50% of the time, its efficacy should be clear in 6 months, after about 150 infections have accumulated in the trial.

Both efforts will pit multiple vaccine candidates head to head. One difference is that Solidarity plans to compare all its vaccines against a shared placebo group, an approach that reduces the number of volunteers the researchers need to recruit and follow. "In the Solidarity trial, the philosophy is we have to make this thing really simple," says Gilbert, who has worked with this effort, too. Solidarity trial sites have the option to do nuanced studies on immunity and other issues, but those are built into Warp Speed's trials. They will do repeated blood draws and nasal or throat swabs to evaluate immune responses and viral levels. The data might help researchers understand why vaccines succeed or how they might affect transmission.

> **"We need to be really careful how we manage expectations."**
> **Wayne Koff**, Human Vaccines Project

In addition to Solidarity, WHO is helping the Access to COVID-19 Tools (ACT) Accelerator, another global effort, which may stage its own vaccine efficacy trials if companies do not want to participate in Solidarity. "Companies may or may not be very enthusiastic about head-to-head comparisons," explains Soumya Swaminathan, WHO's chief scientist and top liaison to the ACT Accelerator. And the ACT Accelerator has pockets as deep as Warp Speed: Countries and philanthropies in May pledged $8 billion, with a commitment that it would equitably distribute any proven COVID-19 products—vaccines, treatments, diagnostics—to rich and poor alike.

Buchbinder is impressed by the speed at which these massive efforts have gotten underway. "It's unlike any other research I've undertaken," she says. But she and others are careful to temper expectations. Even though she will oversee a Warp Speed trial site, for example, she doubts the U.S. effort will meet Trump's goal of having a proven vaccine by October. Koff agrees; the failures of so many HIV vaccine trials have sobered him, he says. "We need to be really careful how we manage expectations," he concludes. ■

COVID-19

Can phone apps slow the spread of the coronavirus?

Digital contact tracing is growing—and facing its first real-world tests

By **Kelly Servick**

Health departments around the world are betting on technology to help stem the stealthy spread of the coronavirus: cellphone apps that aim to identify and alert those who recently came into contact with an infected person. By encouraging those potentially exposed to COVID-19 to self-isolate, the thinking goes, a phone app could swiftly cut off chains of transmission. Dozens of local and national governments have launched official apps or are developing them.

"It's very appealing—that you have an app that does all this work," says Hannah Clapham, an epidemiologist at the National University of Singapore. But she and others warn that an app can't replace human contact tracers. "I worry that we think it's going to save us."

So far, only epidemiological models suggest apps can change a pandemic's course. Ensuring that an app detects risky contacts without overwhelming users with false alarms is one challenge; getting enough people to download an app is another. As health officials weigh competing apps and prepare pitches to privacy-conscious citizens, epidemiologists, engineers, and behavioral scientists are considering how to put an app to the test.

People can apparently transmit the coronavirus for days before they develop symptoms, so by the time health departments learn of a case, they have precious little time before infected contacts start to spread the virus. "You have a couple days to chase people down," says C. Jason Wang, a health policy researcher at Stanford University who works with health departments on COVID-19. And traditional contact tracing—interviewing the infected person, tracking down the recent contacts they can recall, and telling those people to self-isolate—is time consuming.

In contrast, phones could detect when two

Downloaded from http://science.sciencemag.org/ on July 13, 2020

Published by AAAS

Millions of people have downloaded apps designed to alert them to coronavirus exposure.

PHOTO: MINZAYAR OO/PANOS PICTURES/PANOS PICTURES/REDUX

users are close enough to share the virus, and an app could alert one when the other gets sick—even if those people are strangers who happened to sit in adjacent subway seats. "The technology response is absolutely necessary," Wang says, "and it needs to be fast."

Some Asian countries that faced early outbreaks have already put smartphone tracking to use. South Korea collects GPS data to publicize recent paths of newly infected people, and last week announced it would require facilities including bars and night clubs to collect identifying information from visitors' phones to share with health authorities. The Chinese government has combined location data with other surveillance to restrict citizens' movement if they may be infected. GPS location tracking is also a component of apps in India, Iceland, and U.S. states including North Dakota and Utah. But GPS technology isn't precise enough to gauge short distances between two phones. And widespread location tracking raises privacy concerns.

Many governments are instead developing apps that rely on short-range Bluetooth radio signals, which can detect close encounters but don't track movements. Each phone broadcasts an ID number to nearby phones that record such "handshakes." If a user has symptoms or tests positive, the app can trigger alerts to those recent contacts.

To be effective, these apps will first have to reliably estimate distance, based on the strength of the received signals. In a test of Singapore's TraceTogether app, computer scientists Douglas Leith and Stephen Farrell at Trinity College Dublin found flaws in those estimates. The study, posted online but not peer reviewed, showed that when people sat across a table, the received signals were much weaker if their phones were in their pockets versus on the table. Sometimes the signal increased as people moved farther apart—perhaps because of reflection off metal surfaces such as supermarket shelves. The imprecision could lead to both missed encounters and false alarms, Leith warns.

Imperfect Bluetooth readouts can still help if interpreted conservatively, says Marcel Salathé, an epidemiologist at the Swiss Federal Institute of Technology Lausanne. He advises the Swiss government on an app now in a pilot phase that aims to detect when someone comes within 2 meters of an infected person for at least 15 minutes. He thinks his team can tune the system so that "if somebody gets an exposure notification, we will feel damn sure it's actually been a contact."

An app can only catch interactions between people who have installed it. Singapore, which launched TraceTogether on 20 March, says roughly one-third of the population has signed on. But uptake of an app "needs to be almost improbably high to really capture all of the contacts that might be relevant," says Allison Black, a genetic epidemiologist at the University of Washington, Seattle, and the Fred Hutchinson Cancer Research Center.

Recent modeling by infectious disease epidemiologist Christophe Fraser and colleagues at the University of Oxford predicted that if about 56% of a population used an app, it alone could reduce the virus' reproduction number—how many people catch the virus from each infected person—enough to control the outbreak. But Salathé says that, even at relatively low levels of uptake, an app could still prevent infection and save lives: "As soon as you have double digits, I think the effect is already quite substantial."

An app's impact depends on whether people stay home when it tells them to. Human contact tracers connect a potentially infected person to social supports that help them self-quarantine—for example by arranging grocery delivery or even a hotel stay. But because most apps keep users anonymous, health officials won't automatically know who gets an alert or what they need to stay home, Wang says. And users might not dutifully check in with health officials when they get an alert. "That's too optimistic," he says. "We tell people to stay at home; they go to the beach."

To test whether an app accurately flags risky encounters, health officials want to know the proportion of identified contacts who end up sick—the "secondary attack rate." As a rule of thumb, if an app's attack rate matches or exceeds that of traditional contact tracing, "we know the app is doing a really good job," Salathé says.

Some app designs allow health officials access to anonymized ID codes of infected users and all their contacts. That means officials can calculate the secondary attack rate and fine-tune the app by checking how many notified users later report symptoms or a positive test through the app. In Norway, which launched such a centralized app in April, municipalities will compare how many, and how quickly, contacts are identified by the app versus through traditional contact tracing, says Emily MacDonald, an epidemiologist at the Norwegian Institute of Public Health.

Other apps, including Switzerland's and one released this week in Germany, are decentralized, meaning data about interactions stay on a phone. Privacy advocates favor this design, and Google and Apple encourage it. Last month, they released technology to support Bluetooth tracking, but only for apps with a decentralized design. With these apps, health departments will know about potentially infected users only if they report getting an alert. Critics say that design could make it harder to evaluate the app's performance.

The ultimate test, some researchers say, is a randomized trial to gauge whether using an app brings down rates of infection. But such trials "would be very costly and difficult," says Oxford behavioral economist Johannes Abeler. Because COVID-19 is relatively rare, such trials would have to be large. And to gauge effectiveness, researchers would have to factor in the proportion of a participant's contacts who had downloaded the app—which might be nearly impossible to know.

Another possibility is to compare changes in infection rates between geographic areas or demographic groups with different levels of app use, suggests Rosalind Eggo, an epidemiologist at the London School of Hygiene & Tropical Medicine who hopes to study the impact of apps as data accumulate. "We have a lot of technology that can help us here," she says. "There's quite a lot of people saying, 'Oh, it won't work.' I think we need to try." ■

Published by AAAS

COVID-19

U.S. academic research funding stays healthy despite pandemic

Labs reopening even as campuses struggle

By **Jeffrey Mervis**

The coronavirus pandemic has put U.S. higher education in a deep financial hole, triggering hiring freezes, benefit cuts, and a slew of other cost-cutting measures designed to offset huge projected revenue losses. Yet amid the uncertainty about state aid, the supply of foreign students, and other intangibles, one revenue stream has remained healthy: the money universities receive from others—especially the federal government—to carry out research. That fact looms large in any effort to forecast the long-term impact of COVID-19 on U.S. academic research.

In 2018, the federal government provided 53% of the $79 billion spent on research on U.S. campuses, and federal spending has actually increased during the pandemic. This spring, Congress included an additional $3.6 billion for research related to COVID-19 in a string of economic relief packages. Academic leaders are hoping legislators will add to that total in the next stimulus package.

Such developments have put research administrators like Chris Cramer in an enviable position during ongoing budget meetings with other senior academic officers. "These days, I'm the only one who gets to say that, even with large error bars, none of my projected numbers are in the red," says Cramer, who oversees the University of Minnesota's $870 million portfolio of sponsored research.

In part, that's because the pandemic has had little impact on the U.S. government's machinery for awarding grants, even as the number of U.S. COVID-19 cases has surpassed 2 million. At the National Institutes of Health (NIH), which provides about two-thirds of all federal funding for academic research, "We went to 100% virtual peer review [of grant proposals] practically overnight, and it's gone remarkably well," says Michael Lauer, the agency's director of extramural research.

A temporary change in the rules for managing federal grants has also helped buoy academic research. For the past 3 months, institutions have been allowed to continue to pay the salaries of researchers on grants even if their labs have been shuttered.

The 19 March directive, part of a broader strategy to save jobs, reflects the fact that federal grants have always covered much more than the cost of just running experiments or collecting field data. Grantees are also paid for analyzing data, writing papers, supervising students and postdocs, and submitting proposals for new and follow-up studies—work that can all be done from home. (The guidelines were set to expire as *Science* went to press, but could be extended.)

At Johns Hopkins University, which carries out more sponsored research—some $2.5 billion—than any other university in the country, President Ronald Daniels warned this spring of a possible $475 million budget shortfall next year. The bulk of it comes from the suspension of elective procedures at Johns Hopkins's vast network of hospitals and clinics and will be borne by physicians.

The impact on research will be comparatively small, his analysis suggests. A 1-year hiring freeze will save $40 million, Daniels predicted. But the unceasing competition among top research institutions to attract talent and funding all but guarantees that there will be exceptions. "The university must emerge from this crisis stronger, and making strategically important faculty hires is crucial to that end," a university spokesperson says.

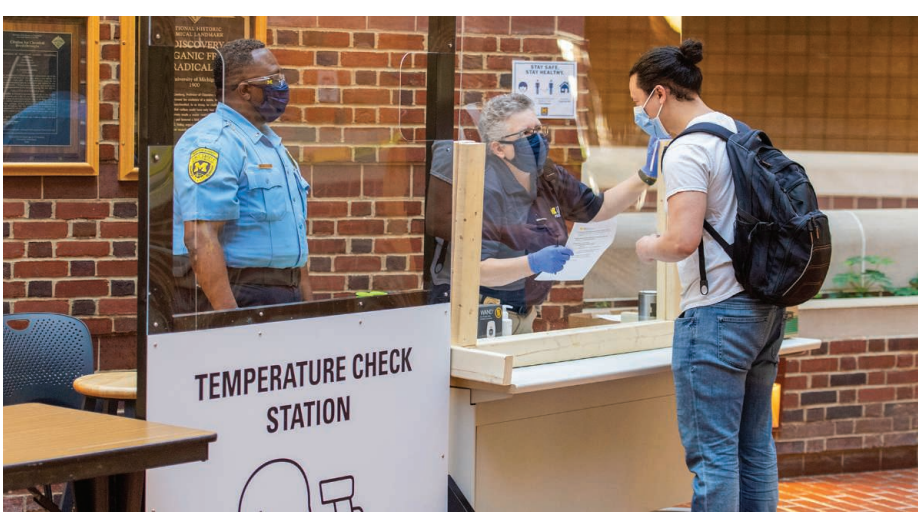
Although many institutions are still figuring out what the fall semester will look like for undergraduates, almost all have already begun to reopen their labs. This week the University of Michigan, which ranks second to Johns Hopkins in sponsored research spending, entered the third phase of a four-step plan to reopen, by the end of the month, all 50 of its research buildings, housing some 5000 labs. "There's a bimodal distribution" of faculty attitudes toward the reopening, says Rebecca Cunningham, the school's vice president for research. "Some are trampling down the doors, and others have serious concerns."

At most institutions, some types of research are still off-limits, including work involving human subjects and most field studies. Last week in *Science*, Cunningham and five other senior administrators from major research universities offered guidance to institutions on how to reopen labs safely. Their suggestions include performing health checks at entrances, controlling access, reducing density, revising work schedules, and requiring personal protective equipment and deep cleaning. "It's a new way of working," Cunningham says, "like with grocery shopping. But people understand that it's necessary."

Research will not emerge from the shutdown unscathed. NIH estimates $10 billion has simply "disappeared" from its budget as a result of lost productivity, Director Francis Collins told a Senate panel last month. Universities have also had to shoulder the unanticipated costs of shutting down and restarting research on their campuses.

A 330-member coalition of research advocates wants Congress to add $31 billion to NIH's current $41 billion budget to make up for such losses and to address infrastructure needs at NIH and around the country. Many of the same organizations are also seeking $46 billion in direct payments to universities to recoup lost revenues.

University officials hope the prominent role their scientists are playing in fighting the pandemic will help persuade lawmakers to provide the additional funding. "When have research universities ever been more in the news?" Cramer asks. "We're on the front pages every day." ■



University of Michigan researchers have their temperatures taken as part of new health and safety procedures.

PHOTO: ROGER HART/MICHIGAN PHOTOGRAPHY

Published by AAAS

ARCHAEOLOGY

Incest in ancient Ireland suggests an elite ruled early farmers

DNA from massive Newgrange tomb reveals a practice linked to royalty around the world

By **Andrew Curry**

Twenty-five kilometers north of Dublin, a masterpiece of Stone Age engineering rises from the hills: a circular structure 12 meters high, almost the area of a U.S. football field, and made up of more than 200,000 tons of earth and stone. Some of the first farmers to arrive in Ireland erected this monument, called Newgrange, nearly 1000 years before Stonehenge or Egypt's first pyramids were built. Archaeologists have assumed it was a ceremonial site and communal tomb—an expression of an egalitarian society.

Now, DNA from a middle-aged man buried in 3200 B.C.E. at the center of this mighty mound suggests otherwise. His genes indicate he had parents so closely related they must have been brother and sister or parent and child.

Across cultures, incest is almost always taboo—except in inbred royal families. Its genetic traces at Newgrange suggest social hierarchy took hold in Ireland earlier than thought, according to a new study. "Maybe we've been arguing too far that [these people were] egalitarian," says Jessica Smyth, an archaeologist at University College Dublin who was not part of the team.

The newly sequenced genomes from Newgrange and other Irish tombs are part of a wider re-evaluation of the Neolithic era, which is marked by the advent of agriculture. Over the past decade, researchers have used ancient DNA to track a slow-motion, 5000-year expansion of ancient farmers from Anatolia across Europe. The Neolithic settlers who arrived in Ireland around 3700 B.C.E. were the westernmost limit of that expansion.

Most Irish Neolithic settlements are small scale, with houses of roughly equal size. As seen in Neolithic graves across Europe, their burials show little sign of hierarchy. Even in major monuments like Newgrange, human remains were jumbled together, as if in a communal tomb. "Archaeologists [have] argued for a long time for a more egalitarian Neolithic," says co-author Thomas Kador, an archaeologist at University College London.

Newgrange is pierced by a passage that leads to a central chamber; its entrance is oriented so a ray of sunlight illuminates the chamber at dawn on the shortest day of the year. "It's clearly a place of public ritual and must have taken a lot of manpower to construct," says geneticist Lara Cassidy of Trinity College Dublin. Hundreds of such passage tomb monuments are found across Ireland.

Most of the bodies in those tombs were cremated. But at the heart of Newgrange, excavators in the 1970s found the unburnt bones of one man, labeled NG10, in a niche decorated with elaborately carved stones. Cassidy and her team were able to extract DNA from that skeleton.

Comparing NG10's DNA and that of other Neolithic burials with DNA from people living on the island centuries earlier shows Neolithic farmers arrived in Ireland as part of a mass migration, and soon swamped or eliminated the genetic legacy of earlier hunter-gatherers, says geneticist Daniel Bradley, a co-author also at Trinity College Dublin.

NG10's DNA also reveals his unusual parentage. In a paper this week in *Nature*, Cassidy's team draws on parallels in the historical record to argue that the son of an incestuous union buried in such a prominent tomb points to a hereditary ruling class. "Matings like that are taboo pretty much universally, with very few exceptions," she says.

Those exceptions include Egyptian pharaohs, who were considered deities who needed to marry each other. Royal siblings in Hawaii and the Incan empire were also known to marry, concentrating power in one family. "I believe we're seeing a similar social dynamic at play among colonists of Neolithic Ireland," Cassidy says.

Additional DNA from more than 40 people buried at other Neolithic sites, including three passage tombs, supports the existence of a close-knit elite. People buried in passage tomb sites were more closely related to each other than to people buried in other types of tombs, even though the passage tombs were separated by hundreds of kilometers and spanned more than 500 years. Some individuals in the far-flung passage tombs could have been second or third cousins or great-great-great-great-grandparent and child.

Chemical isotopes in their bones show the people in the passage tomb burials ate more meat and animal products than their contemporaries. The burials also include women and children, suggesting social status was inherited rather than won in a single lifetime. "We're talking about something that could be inherited," Kador says. "A small [related] elite called the shots, like in Egypt."

But other archaeologists are cautious. "To go from [NG10] to saying these are proto-state societies where you have a godlike elite is pushing it a bit far," says University of Manchester archaeologist Julian Thomas. "It's one guy." He notes Newgrange was a burial place for almost 1000 years, too long to make generalizations from a single burial.

One way to settle the debate is to look at similar passage tombs built on the Orkney Islands and in Wales and France. "The question is whether this arose in Ireland or whether they were importing existing social structures into the island," Cassidy says. "It's going to be very exciting to see if this is a pattern we see in other areas." ■

Early Irish royalty may have been buried in Newgrange tomb, shown at dawn on the winter solstice.

PHOTO: KEN WILLIAMS/SHADOWSANDSTONE.COM

Andrew Curry is a journalist in Berlin.

Published by AAAS

In late 2019, reporters visited the Arak Heavy Water Complex, a potential source of plutonium.

NUCLEAR PROLIFERATION

New tensions dim hopes for salvaging Iran nuclear deal

Monitoring agency seeks clarity about clandestine R&D

By **Richard Stone**

Since U.S. President Donald Trump pulled the United States out of the Iran nuclear deal 2 years ago, other signatories have tried to salvage the agreement in hopes of constraining Iran's ability to resurrect a nuclear weapons program. They haven't given up yet. And although Iran has resumed activities proscribed by the deal, including stepping up uranium enrichment, it has kept the door open to a reversal. But a series of disputes has continued to fray the agreement.

This week, questions about Iran's past clandestine atomic research program—including the disappearance of a missing uranium disk that could be used in a bomb, and illicit explosives testing—took center stage during a meeting of the governing board of the International Atomic Energy Agency (IAEA). And recent moves by the Trump administration have threatened to derail the conversion of Iranian nuclear facilities to reduce the risk they could contribute to a weapons program.

The Joint Comprehensive Plan of Action (JCPOA), as the 2015 deal is called, offered Iran relief from economic sanctions in exchange for dismantling large pieces of its nuclear program. Although experts generally agree the JCPOA worked, the Trump administration withdrew in May 2018, arguing it didn't go far enough. New U.S. sanctions have since battered Iran's economy.

"It hasn't been easy to watch the deal get dismantled," says a European diplomat who requested anonymity because talks with Iran are at a sensitive stage. Yet, "So far, Iran has been notably measured in building up its nuclear capabilities," Christopher Ford, assistant secretary for international security and nonproliferation at the U.S. State Department, recently told journalists.

An IAEA report shared with the governing board last week said Iran has been enriching uranium hexafluoride gas to 4.5% of the fissile isotope uranium-235 (U-235) over the past year. By 20 May, it had stockpiled 1572 kilograms of enriched uranium, ostensibly for use in civilian reactors. Nuclear bombs require enrichment levels exceeding 90% of U-235. "They could have made a lot more uranium at a higher enrichment level, but they haven't," says Richard Johnson of the nonprofit Nuclear Threat Initiative.

But Iran has greatly ramped up R&D on advanced centrifuges, which could speed up enrichment and reduce the "breakout time" needed to enrich a bomb's worth of weapons-grade uranium to 3 months. The JCPOA ensured the breakout time would exceed 1 year; the Western diplomat estimates it is still more than 6 months.

IAEA and outside analysts have also been poring over a cache of documents Israeli agents spirited out of Iran in early 2018. The nuclear archive, as it's called, has yielded fresh insights into Iran's past R&D on nuclear weapons and its plans for underground testing that IAEA says call into question the "correctness and completeness" of declarations Iran made in 2003, when it agreed to come clean on its nuclear program and permit inspectors broad access to sites.

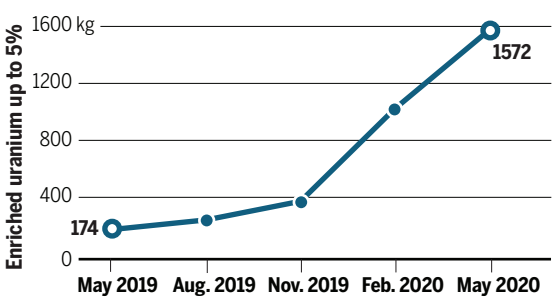
Iran had hoped the deal would end questions about its shuttered bomb effort. But IAEA now wants to know the whereabouts of a disk of uranium metal that could be used to generate neutrons for triggering fission in a bomb's U-235 core. Evidence suggests it was housed at a site called Lavizan-Shian in Tehran, which Iran razed and sanitized in 2003 and 2004. In January, IAEA asked Iran to give its inspectors access to two unnamed sites to verify the absence of undeclared nuclear material and activities. According to a report last week from the nonprofit Institute for Science and International Security, one site appears to be a testing range for high explosives near Abadeh, in central Iran. During a September 2019 press conference, Israeli Prime Minister Benjamin Netanyahu showed satellite images indicating that site was razed only in July 2019.

In a 28 January letter to IAEA, Iran rebuffed the agency, declaring it "will not recognize any allegation on past activities and does not consider itself obliged to respond to such allegations." The dispute has raised fresh questions about Iran's intentions and gives the United States another cudgel to try to persuade the United Nations Security Council to reimpose nuclear related sanctions.

Other components of the JCPOA are also foundering. Before the agreement, Iran was building a heavy water research reactor in Arak that would accumulate several kilograms of plutonium a year in spent fuel—enough for one or two bombs. The deal mandated a redesign of the Arak reactor to sharply curtail generation of plutonium. Even after walking away from the JCPOA,

An enriching pursuit

Iran has revved up uranium enrichment, decreasing the time needed to produce a bomb's worth of uranium-235.



CREDITS: (PHOTO) SALAMPIX/ABACA/SIPA USA/AP IMAGES; (GRAPH) M. ENSERINK/*SCIENCE*; (DATA) DAVID ALBRIGHT, SARAH BURKHARD, AND ANDREA STRICKER/INSTITUTE FOR SCIENCE AND INTERNATIONAL SECURITY

Published by AAAS

the United States had supported the redesign by waiving sanctions on other countries taking part in the work—until May, when the State Department decided to let the waivers expire as of 27 July. The move perplexed observers. "I cannot stress enough how bizarre it is to me that we demanded that the Iranians convert the reactor—and now we insist they must not convert it," says Jeffrey Lewis of the Middlebury Institute of International Studies at Monterey.

If the Arak redesign falls through, Iran could claim that Western powers are out of compliance with the JCPOA, the diplomat says. Work will continue at Arak, which Iran has renamed Khondab, says Ali Akbar Salehi, president of the Atomic Energy Organization of Iran. "Soon the international community will witness our new achievements at the Khondab Research Reactor," Salehi told *Science*. "Although sanctions impose some constraints, it invigorates us." He did not provide details, and analysts are not sure what Iran's plans are for the reactor.

U.S. withdrawal from the JCPOA has also scuttled a grand plan to turn an underground facility near the holy city of Qom into an international research center. Under the deal, Iran mothballed 700 uranium enrichment centrifuges at the Fordow site and worked with Russia to convert 328 others to producing isotopes for medicine. JCPOA negotiators floated other ideas, including installing a particle accelerator in the cramped space. But late last year, Iran resumed uranium enrichment in one Fordow hall. "That created a bit of a conundrum," says Johnson, who was involved in the deal's implementation during the Obama administration. Even minute traces of uranium would contaminate the medical isotope centrifuges. Worsening matters, a few weeks later, the United States canceled a sanctions waiver for the medical isotope work, prompting Russia to back away. Iran will continue to operate a dozen Russian-modified centrifuges on its own.

The failure to convert Fordow to a civilian research center "is a missed opportunity," says Andrea Stricker, a nonproliferation analyst at the nonprofit Foundation for Defense of Democracies.

Despite the nuclear deal's slow-motion collapse, observers don't expect Iran to open up the throttle on its program—at least not before the U.S. elections in November. If Democratic presidential candidate Joe Biden wins, the next administration "may try to resurrect some form of the JCPOA," Stricker says. "And the Iranians would probably want to test and see what they can get." ■

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

SCIENTIFIC COMMUNITY

NIH requires disclosure about sexual harassment by grantees

Changes aimed in part at making sure institutions don't "pass the harasser"

By **Jocelyn Kaiser**

The National Institutes of Health (NIH) is tightening grant rules to demand institutions tell it about sexual harassment by researchers it supports. Starting 12 June for new awards, NIH will expect institutions it funds to report when an investigator is removed from a grant because of harassment findings or allegations. NIH also wants to know when an investigator moves their grant to another institution because of sexual misconduct issues.

Director Francis Collins and other officials announced the new reporting guidance in an online editorial in *Science* last week. Together, the requirements will "close two important gaps" in the agency's policies, says NIH Associate Director for Science Policy Carrie Wolinetz, and should prevent cases in which institutions "pass the harasser" without the agency's knowledge. The changes will "further foster a culture whereby sexual harassment and other inappropriate behaviors are not tolerated," NIH officials write.

> **"It's definitely a step in the right direction."**
> **Carol Greider**, Johns Hopkins University

The changes align NIH's policies more closely with those of the National Science Foundation (NSF), although a gap remains. NSF requires institutions to notify the agency within 10 days of sexual misconduct findings—even if there's no change in an investigator's grant status. NIH has said that because it is not an independent agency like NSF, it doesn't have the legal authority to adopt that policy without lengthy formal rulemaking.

But NIH does have authority to ask about changes in grant personnel. Institutions must seek NIH approval for such changes, which can be due to factors including medical leave, job changes, or misconduct investigations. Until now, NIH did not ask the reasons. But the agency recently realized it could fold harassment into grant rules that require "safe and healthful working conditions," Wolinetz says. An 11 June notice says any request to NIH to change the grant's key personnel should "mention as to whether change(s) ... is related to concerns about safety and/or work environments (e.g. due to concerns about harassment, bullying, retaliation, or hostile working conditions)."

If an institution reports that a grant's personnel changed because of allegations or findings of harassment, NIH staff will review the information and decide, for example, whether the investigator can keep the grant under supervision, or the grant should end or be transferred. If a grantee is moving to a new institution because of harassment concerns, the agency plans to inform the new institution about the investigator's record.

The changes "will hopefully capture the vast majority of cases," Wolinetz says.

"I'm very pleased. It's definitely a step in the right direction," says Johns Hopkins University biologist and Nobel laureate Carol Greider, a member of a working group of NIH's Advisory Committee to the Director (ACD) that in December 2019 advised NIH to make the policy changes. But Greider worries about loopholes.

Contrary to NSF's rules, for example, institutions won't have to report findings of sexual harassment by an investigator to NIH if the investigator's status on a grant is unchanged—a step that was recommended by the working group. And such situations appear to be "a worrisome trend," Wolinetz noted last week at an ACD meeting. The growing number of sexual harassment cases voluntarily reported to NIH has made it clear that universities sometimes remove an investigator from training and supervisory duties without transferring their research grants. It seems that "Institutions are protecting their rainmakers," Wolinetz said.

Francis Cuss, a retired Bristol Myers Squibb official who co-chaired the working group, suggested NIH's policies don't go far enough and "may need ... more teeth."

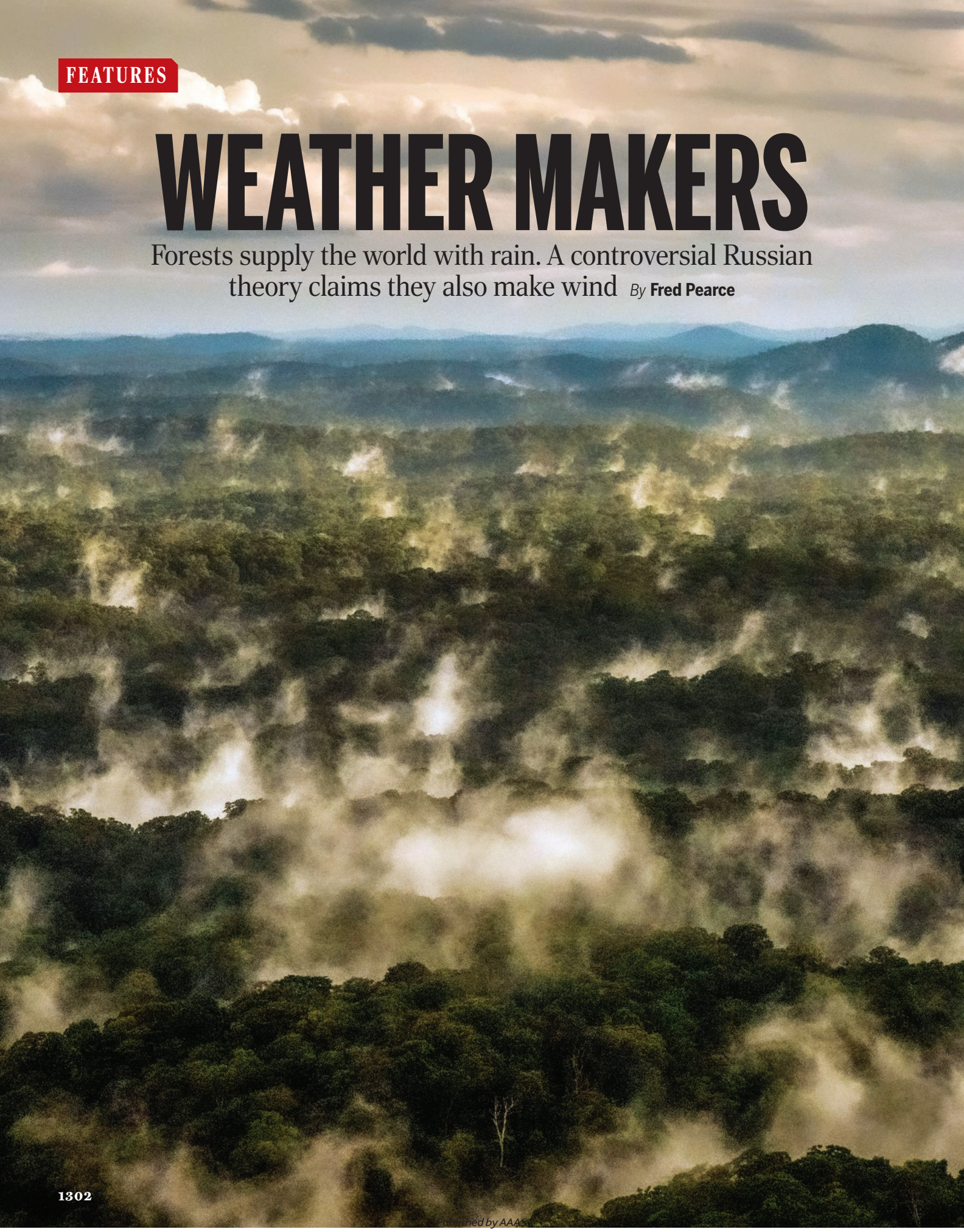
Columbia University virologist Angela Rasmussen, another working group member, agrees. The new policy "is a necessary first step in a much longer journey," she says. ■

Published by AAAS

WEATHER MAKERS

Forests supply the world with rain. A controversial Russian theory claims they also make wind

By **Fred Pearce**



Published by AAAS

Every summer, as the days get long, Anastassia Makarieva leaves her lab in St. Petersburg for a vacation in the vast forests of northern Russia. The nuclear physicist camps on the shores of the White Sea, amid spruce and pine, and kayaks along the region's wide rivers, taking notes on nature and the weather. "The forests are a big part of my inner life," she says. In the 25 years she has made her annual pilgrimage north, they have become a big part of her professional life, too.

For more than a decade, Makarieva has championed a theory, developed with Victor Gorshkov, her mentor and colleague at the Petersburg Nuclear Physics Institute (PNPI), on how Russia's boreal forests, the largest expanse of trees on Earth, regulate the climate of northern Asia. It is simple physics with far-reaching consequences, describing how water vapor exhaled by trees drives winds: winds that cross the continent, taking moist air from Europe, through Siberia, and on into Mongolia and China; winds that deliver rains that keep the giant rivers of eastern Siberia flowing; winds that water China's northern plain, the breadbasket of the most populous nation on Earth.

With their ability to soak up carbon dioxide and breathe out oxygen, the world's great forests are often referred to as the planet's lungs. But Makarieva and Gorshkov, who died last year, say they are its beating heart, too. "Forests are complex self-sustaining rainmaking systems, and the major driver of atmospheric circulation on Earth," Makarieva says. They recycle vast amounts of moisture into the air and, in the process, also whip up winds that pump that water around the world. The first part of that idea—forests as rainmakers—originated with other scientists and is increasingly appreciated by water resource managers in a world of rampant deforestation. But the second part, a theory Makarieva calls the biotic pump, is far more controversial.

The theoretical foundation of the work has been published, albeit in lesser known journals, and Makarieva has received support from a small coterie of colleagues. But the biotic pump has faced a head wind of criticism, especially from climate modelers, some of whom say its effects are negligible and dismiss the idea completely. The dispute has made Makarieva an outsider: a theoretical physicist in a world of modelers, a Russian in a field led by Western scientists, and a woman in a field dominated by men.

Half of the Amazon's rain comes from the forest's own moisture. Could it also make winds that ferry rain across continents?

Yet, if correct, the idea could help explain why, despite their distance from the oceans, the remote interiors of forested continents receive as much rain as the coasts—and why the interiors of unforested continents tend to be arid. It also implies that forests from the Russian taiga to the Amazon rainforest don't just grow where the weather is right. They also make the weather. "All I have learned so far suggests to me that the biotic pump is correct," says Douglas Sheil, a forest ecologist at the Norwegian University of Life Sciences. With the future of the world's forests in doubt, "Even if we thought the theory had only a small chance of being true, it would be profoundly important to know one way or the other."

Anastassia Makarieva and Victor Gorshkov developed the biotic pump theory at a nuclear physics institute.

PHOTOS: ANASTASSIA MAKARIEVA; (OPPOSITE PAGE) BRUSINI AURÉLIEN/HEMIS/ALAMY STOCK PHOTO

Many meteorology textbooks still teach a caricature of the water cycle, with ocean evaporation responsible for most of the atmospheric moisture that condenses in clouds and falls as rain. The picture ignores the role of vegetation and, in particular, trees, which act like giant water fountains. Their roots capture water from the soil for photosynthesis, and microscopic pores in leaves release unused water as vapor into the air. The process, the arboreal equivalent of sweating, is known as transpiration. In this way, a single mature tree can release hundreds of liters of water a day. With its foliage offering abundant surface area for the exchange, a forest can often deliver more moisture to the air than evaporation from a water body of the same size.

The importance of this recycled moisture for nourishing rains was largely disregarded until 1979, when Brazilian meteorologist Eneas Salati reported studies of the isotopic composition of rainwater sampled from the Amazon Basin. Water recycled by transpiration contains more molecules with the heavy oxygen-18 isotope than water evaporated from the ocean. Salati used this fact to show that half of the rainfall over the Amazon came from the transpiration of the forest itself.

By this time, meteorologists were tracking an atmospheric jet above the forest, at a height of about 1.5 kilometers. Known as the South American Low-Level Jet, the winds blow east to west across the Amazon, about as fast as a racing bike, before the Andes Mountains divert them south. Salati and others surmised the jet carried much of the transpired moisture, and dubbed it a "flying river." The Amazon flying river is now reckoned to carry as much water as the giant terrestrial river below it, says Antonio Nobre, a climate researcher at Brazil's National Institute for Space Research.

For some years, flying rivers were thought to be limited to the Amazon. In the 1990s, Hubert Savenije, a hydrologist at the Delft University of Technology, began to study moisture recycling in West Africa. Using a hydrological model based on weather data, he found that, as one moved inland from the coast, the proportion of the rainfall that came from forests grew,

Published by AAAS

reaching 90% in the interior. The finding helped explain why the interior Sahel region became dryer as coastal forests disappeared over the past half-century.

One of Savenije's students, Ruud van der Ent, took the idea further, creating a global model of airborne moisture flow. He combined observational data on rainfall, humidity, wind speed, and temperature with theoretical estimates of evaporation and transpiration to create the first model of moisture flow at scales larger than river basins.

In 2010, van der Ent and his colleagues reported the model's conclusion: Globally, 40% of all precipitation comes from the land rather than the ocean. Often it is more. The Amazon's flying river provides 70% of the rain falling in the Río de la Plata Basin, which stretches across southeastern South America. Van der Ent was most surprised to find that China gets 80% of its water from the west, mostly Atlantic moisture recycled by the boreal forests of Scandinavia and Russia. The journey involves several stages—cycles of transpiration followed by downwind rain and subsequent transpiration—and takes 6 months or more. "It contradicted previous knowledge that you learn in high school," he says. "China is next to an ocean, the Pacific, yet most of its rainfall is moisture recycled from land far to the west."

IF MAKARIEVA IS CORRECT, the forests supply not just the moisture, but the winds that carry it.

For a quarter-century, she worked with Gorshkov, initially as his pupil, at PNPI—part of Russia's foremost civil and military nuclear research agency, the Kurchatov Institute. They were mavericks from the start, studying ecology in a place full of physicists who use neutron beams from nuclear reactors to study materials. As theorists, she says, they had "exceptional freedom of research and thought," pursuing atmospheric physics wherever it took them. "Victor taught me: Do not be afraid of anything," she says.

In 2007, in *Hydrology and Earth System Sciences*, they first outlined their vision for the biotic pump. It was provocative from the outset because it contradicted a longstanding tenet of meteorology: that winds are driven largely by the differential heating of the atmosphere. When warm air rises, it lowers the air pressure below it, in effect creating space at the surface into which air moves. In summer, for example, land surfaces tend to heat faster and draw in moist breezes from the cooler ocean.

Makarieva and Gorshkov argued that a second process can sometimes dominate. When water vapor from forests condenses into clouds, a gas becomes a liquid that occupies less volume. That reduces air pressure, and draws in air horizontally from areas with less condensation. In practice, it means condensation above coastal forests turbocharges sea breezes, sucking moist air inland where it will eventually condense and fall as rain. If the forests continue inland, the cycle can continue, maintaining moist winds for thousands of kilometers.

The theory inverts traditional thinking: It is not atmospheric circulation that drives the hydrological cycle, but the hydrological cycle that drives the mass circulation of air.

Sheil, who became a supporter of the theory more than a decade ago, thinks of it as an embellishment of the flying river idea. "They are not mutually exclusive," he says. "The pump offers an explanation of the power of the rivers." He says the biotic pump could explain the "cold Amazon paradox." From January to June, when the Amazon Basin is colder than the ocean, strong winds blow from the Atlantic to the Amazon—the opposite of what would be expected if they resulted from differential heating. Nobre, another early acolyte, enthuses: "They don't start with data, they start with first principles."

"We need new international hydrological agreements to maintain the forests of source regions."

Lan Wang-Erlandsson, Stockholm University

Even those who doubt the theory agree that forest loss can have far-reaching climatic consequences. Many scientists have argued that deforestation thousands of years ago was to blame for desertification in the Australian Outback and West Africa. The fear is that future deforestation could dry up other regions, for example, tipping parts of the Amazon rainforest to savanna. Agricultural regions of China, the African Sahel, and the Argentine Pampas are also at risk, says Patrick Keys, an atmospheric chemist at Colorado State University, Fort Collins.

In 2018, Keys and his colleagues used a model, similar to van der Ent's, to track the sources of rainfall for 29 global megacities. He found that 19 were highly dependent on distant forests for much of their water supply, including Karachi, Pakistan; Wuhan and Shanghai, China; and New Delhi and Kolkata, India. "Even small changes in precipitation arising from upwind land-use change could have big impacts on the fragility of urban water supplies," he says.

Some modeling even suggests that by removing a moisture source, deforestation could alter weather patterns beyond the paths of flying rivers. Just as El Niño, a shift in currents and winds in the tropical Pacific Ocean, is known to influence weather in faraway places through "teleconnections," so, too, could Amazon deforestation diminish rainfall in the U.S. Midwest and snowpack in the Sierra Nevada, says Roni Avissar, a climatologist at the University of Miami who has modeled such teleconnections. Far-fetched? "Not at all," he says. "We know El Niño can do this, because unlike deforestation, it recurs and we can see the pattern. Both are caused by small changes in temperature and moisture that project into the atmosphere."

Lan Wang-Erlandsson, who researches interactions between land, water, and climate at Stockholm University, says it's time for water resource managers to shift their focus from water and land use within a river basin to land-use changes occurring outside it. "We need new international hydrological agreements to maintain the forests of source regions," she says.

Two years ago, at a meeting of the United Nations Forum on Forests, a high-level policy group on which all governments sit, David Ellison, a land researcher at the University of Bern, presented a case in point: a study showing that as much as 40% of the total rainfall in the Ethiopian highlands, the main source of the Nile, is provided by moisture recycled from the forests of the Congo Basin. Egypt, Sudan, and Ethiopia are negotiating a long-overdue deal on sharing the waters of the Nile. But such an agreement would be worthless if deforestation in the Congo Basin, far from those three nations, dries up the moisture source, Ellison suggested. "Interactions between forests and water have been almost entirely ignored in the management of global freshwater resources."

The biotic pump would raise the stakes even further, with its suggestion that forest loss alters not just moisture sources, but also wind patterns. The theory, if correct, would have "crucial implications for planetary air circulation patterns," Ellison warns, especially those that take moist air inland to continental interiors.

THE THEORY'S SUPPORTERS are a minority. In 2010, Makarieva, Gorshkov, Sheil, Nobre, and Bai-Lian Li, an ecologist at the University of California, Riverside, submitted what was meant to be a landmark description of the biotic pump to *Atmospheric Chemistry and Physics*, a major journal with open peer review. Titled "Where Do Winds Come From?" the paper faced a barrage of criticism online, and it took the journal many months to find two scientists willing to review it. Isaac Held, a meteorologist at Princeton University's Geophysical Fluid Dynamics Laboratory, finally volunteered—and rec-

Published by AAAS

ommended rejection. "This is not a mysterious effect," he says. "It is small and included in some atmospheric models." Critics said the expansion of air from heat released when water vapor condenses counteracts the space-creating effect of condensation. But Makarieva says the two effects are spatially separate, with the warming effect happening aloft, and the pressure drop of condensation occurring closer to the surface, where it generates the biotic wind.

The other reviewer was Judith Curry, then an atmospheric physicist at the Georgia Institute of Technology, who has long had concerns about the atmospheric dynamics at the core of climate models. She felt it was important to publish the paper and says the standoff was "very bad for climate science, which badly needs an infusion from hardcore physicists." After 3 years of debate, the journal's editor overruled Held's recommendation and published the paper, saying it was published "not as an endorsement" but "to promote continuation of the scientific dialogue on the controversial theory [that] may lead to disproof or validation."

Since then, there has been neither validation nor disproof, but largely a standoff. Gavin Schmidt, a climate modeler at Columbia University, says, "It's simply nonsense." The authors' responses to criticisms were "really just mathematics that gave no one any confidence that there was any point in continuing the dialogue." Jose Marengo, a meteorologist in Brazil and head of the National Centre for Monitoring and Warning of Natural Disasters, says: "I think the pump exists, but it's very theoretical right now. The climate model community hasn't embraced it, but the Russians are the best theoreticians in the world, so we need proper field experiments to test it." Yet no one, including Makarieva, has yet proposed clearly what such a test might look like.

For her part, Makarieva is building on the theory, arguing in a series of recent papers that the same mechanism can affect tropical cyclones, which are driven by the heat released when moisture condenses over the ocean. In a 2017 paper in *Atmospheric Research*, she and her colleagues proposed that biotic pumps set up by the forests on land draw moisture-rich air away from the cyclone nurseries. This, she says, might explain why cyclones rarely form in the South Atlantic Ocean: The Amazon and Congo rainforests between them draw so much moisture away that there is too little left to fuel hurricanes.

Kerry Emanuel, a leading hurricane researcher at the Massachusetts Institute of Technology, says the proposed effects "while not negligible are very small." He prefers other explanations for the lack of South Atlantic hurricanes, such as the region's cool waters, which send less moisture into the air, and its strong shearing winds, which disrupt cyclone formation. Makarieva is equally dismissive of the traditionalists, saying some of the existing theories for hurricane intensity "conflict with the laws of thermodynamics." She has another paper on the topic under peer review at the *Journal of the Atmospheric Sciences*. "We are concerned that, despite the editor's encouragement, our work will get rejected once again," she says.

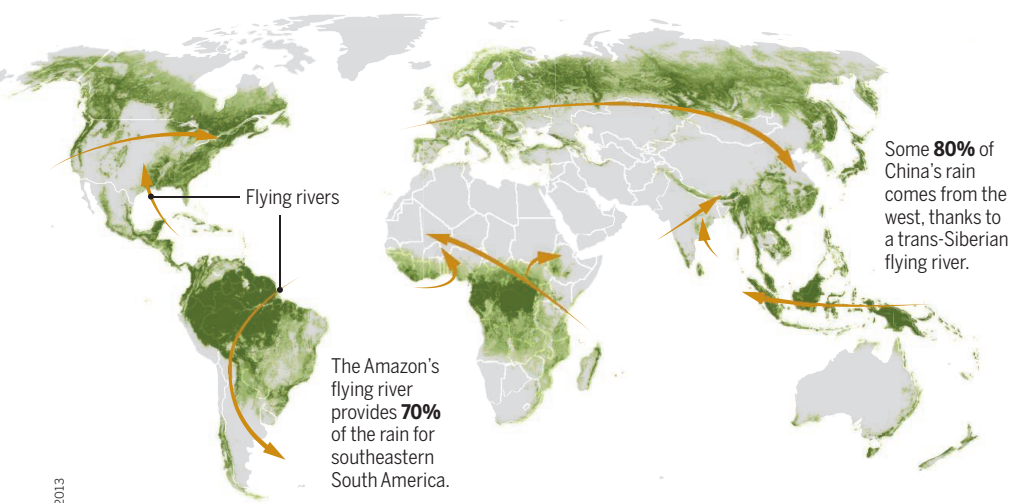
Even if Makarieva's ideas are fringy in the West, they are taking root in Russia. Last year, the government began a public dialogue to revise its forestry laws. Aside from strictly protected areas, Russian forests are open to commercial exploitation, but the government and the Federal Forestry Agency are considering a new designation of "climate protection forests." "Some representatives of our forest department got impressed by the biotic pump and want to introduce a new category," she says. The idea has the backing of the Russian Academy of Sciences. Being part of a consensus rather than the perennial outsider marks a change, Makarieva says.

This summer, the coronavirus lockdown put the kibosh on her annual trip to the northern forests. Back in St. Petersburg, she has settled down to respond to yet another round of objections to her work from anonymous peer reviewers. She insists the pump theory will ultimately prevail. "There is a natural inertia in science," she says. With a dark Russian humor, she invokes the words of the legendary German physicist Max Planck, who is said to have once remarked that science "advances one funeral at a time." ■

Fred Pearce is a journalist in London.

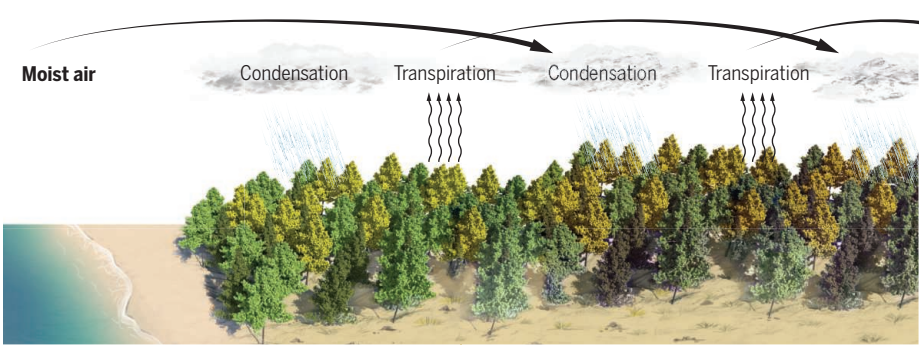
Rain parades

So-called flying rivers are prevailing winds that pick up water vapor exhaled by forests and deliver rains to distant water basins. A controversial theory suggests forests themselves drive the winds (bottom).



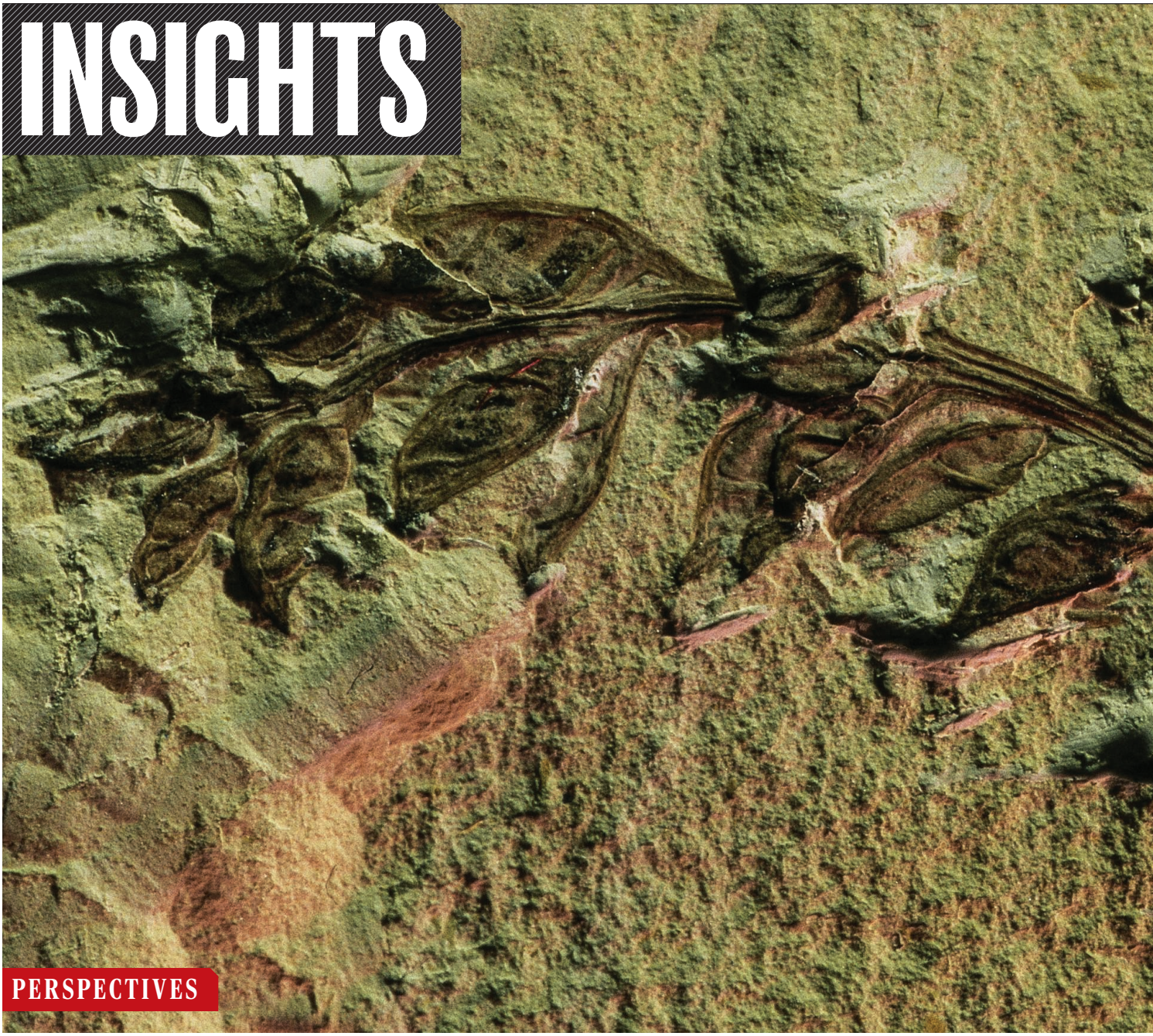
Sowing the wind

The biotic pump theory suggests forests not only make rain, but also wind. When water vapor over coastal forests condenses, it lowers air pressures, creating winds that draw in moist ocean air. Cycles of transpiration and condensation can set up winds that deliver rains thousands of kilometers inland.



CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) HANSEN *ET AL.*, *SCIENCE*, **342** (6160) 2013

Published by AAAS



EVOLUTION

The origins of flowering plants and pollinators

New research raises questions about when flowering plants and their pollinators evolved

By **Casper J. van der Kooi**[1] *and* **Jeff Ollerton**[2]

For more than a century there has been a fascination with the surprisingly rapid rise and early diversity of flowering plants (angiosperms). Darwin described the seemingly explosive diversification of angiosperms as an "abominable mystery," and debates continue about the origin and processes driving angiosperm speciation. Dating the origin of angiosperms was traditionally the prerogative of paleobotanists who read the fossil record of plants, but with DNA sequencing becoming increasingly sophisticated, molecular dating methods have come to the table. Many angiosperm fossils can be dated to the Early Cretaceous (~135 million years ago), which has led paleobotanists to reason that they originated during that era. It is now increasingly recognized that angiosperms are probably older than the oldest fossils, but how much older remains controversial. When angiosperms originated is key to understanding the origin and evolu-

PHOTO: JONATHAN BLAIR/NATIONAL GEOGRAPHIC IMAGE COLLECTION

Published by AAAS

The aquatic angiosperm *Archaefructus liaoningensis* is one of the earliest fossil angiosperms to have been identified so far.

tion of pollinators, particularly insects such as bees, butterflies, moths, and flies.

Recent reports highlight the disparity of molecular and paleontological time scales and draw conflicting conclusions about the timing of angiosperm diversification (see the figure). On the basis of gene sequences from 2881 chloroplast genomes belonging to species from 85% of living flowering-plant families, time-calibrated using 62 fossils, one study (*1*) dated the origin of angiosperms to the Late Triassic, >200 million years ago. This is ~70 million years (roughly the equivalent of the Jurassic) before the earliest accepted angiosperm fossils. This study further suggests that major radiations (species diversification) occurred in the Late Jurassic and Early Cretaceous, ~165 to 100 million years ago. By contrast, an overview of paleobotanical evidence (*2*) refutes a substantive pre-Cretaceous diversification, with only some specific clades (such as water lilies) perhaps originating during the Late Jurassic. The sequential appearance of different types of fossils and morphological characteristics is proposed to render major earlier diversification events unlikely, supporting previous studies (*3*, *4*). Although the idea that angiosperms arose around the beginning of the Cretaceous may seem hard to reconcile with the rapid increase in morphological diversity observed during that interval, it is not impossible if the Cretaceous radiation occurred rapidly.

Both paleontological records and molecular analyses have their strengths and weaknesses. The strength of fossils is that they can provide information on past form, function, and clade richness, and indirectly provide information on speciation and extinction. Fossils are particularly useful when they harbor intermediate structures or combinations of characters that no longer exist, which can provide insightful examples that help to reconstruct the course of evolutionary events. However, the interpretation of fossils can be subjective and controversial, because important features of these plants may not be preserved and often must be inferred from two-dimensional compressed remains.

The absence of evidence is no evidence of absence, and it is known that the fossil record can be incomplete or biased because some taxa may be less likely to fossilize. For example, specific ecologies or habitats will influence the likelihood of whole-plant fossilization, although pollen is a useful exception because it can generally survive more extreme conditions. Furthermore, anchoring a fossil to a specific time period relies on accurately dating the stratum in which it was found, which can also be problematic, although the error margin caused by this factor is usually small. It is important to keep in mind that there can be a considerable lag between time of origin and the earliest recognizable fossil, because fossils generally appear when a taxon has existed for some time and in relatively high frequencies, a phenomenon known as the Signor-Lipps effect.

Molecular analyses are built on hard-to-estimate variables, such as the distribution of mutation rates across taxa and time. Variation in divergence times—which inevitably occurs in datasets with many species—frequently leads to overestimation of age (*5*, *6*). Indeed, molecular analyses often push origin dates back in time, including the older lineages, but whether this is a methodological error remains unclear.

One of the hallmarks of angiosperms is their relationship with animal pollinators, especially insects. As with plants, the diversification of insects is a field with many uncertainties. The origin of several important orders of flower-visiting insects (e.g., Coleoptera, Diptera, Hymenoptera, and Lepidoptera) lies in the Permian or Triassic (300 to 200 million years ago) with marked periods of diversification in the Cretaceous, which is frequently mentioned to coincide with the main angiosperm radiation (*7*). However, the timing of the origin of flower-visiting insects is debated. For example, for Lepidoptera (butterflies and moths), a Late Triassic radiation has been suggested on the basis of fossil evidence (*8*), but a recent study using transcriptomes covering nearly all Lepidoptera superfamilies dated the origin even further back, during the Carboniferous (~300 million years ago) (*9*). Although butterfly diversification may be triggered more by host plant chemistry than by floral diversity—which need not be correlated—given the importance of butterflies and moths for angiosperm reproduction, their diversification is important in understanding plant-pollinator interactions.

Notwithstanding that the timing of the origin of angiosperms remains debated, if angiosperms arose before the Jurassic, this has profound implications for understanding how insect pollination evolved. There is little doubt that insect pollination accelerated the angiosperm radiation; however, which factor triggered what evolutionary event becomes more complex given the latest findings. It was long considered that wind pollination in early-diverging nonflowering seed plants (gymnosperms) was replaced by animal pollination in angiosperms, and that this switch to animal pollination led to angiosperm diversification, but this seems an oversimplification (*10*).

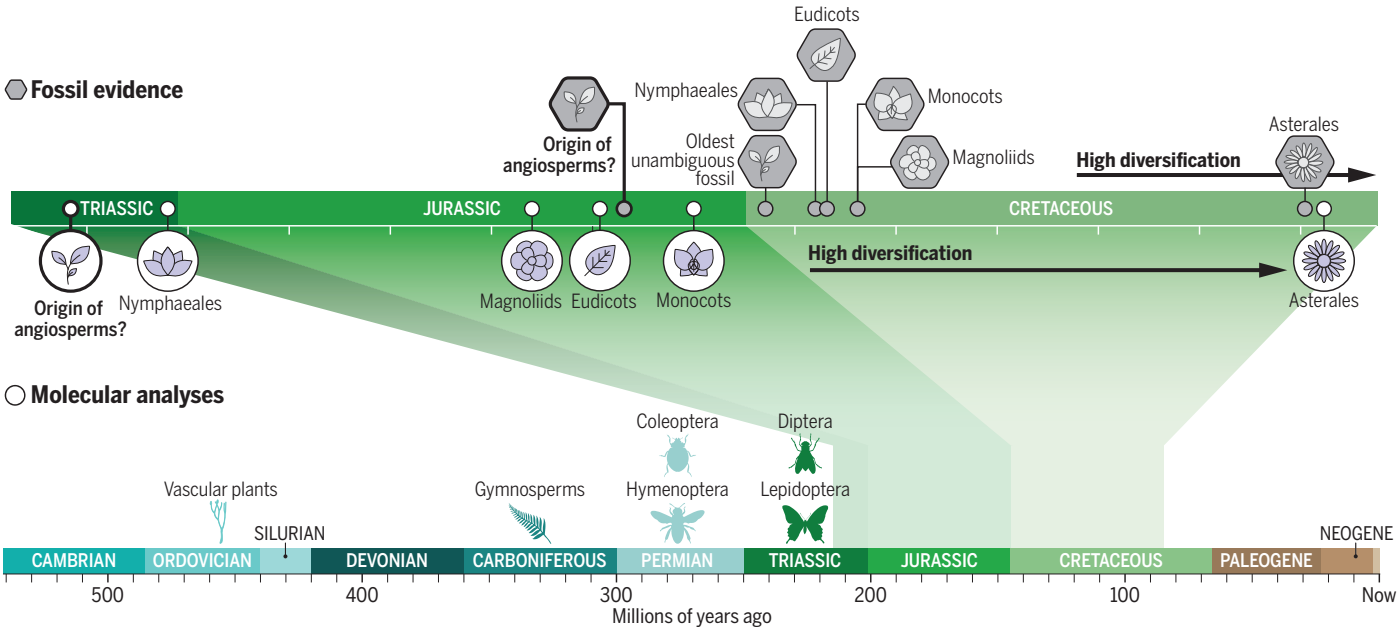
Many now-extinct gymnosperms (e.g., Bennettitales) were insect pollinated, and angiosperms could have evolved either directly from insect-pollinated gymnosperms or from wind-pollinated gymnosperms in such a way that they co-opted insects that were servicing gymnosperms in the same community. Conversely, if the earlier Triassic origin of angiosperms is correct, some gymnosperms may have co-opted insects as pollinators from early angiosperms. It seems unlikely, however, that this latter process was important in

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, NL-9747AG Groningen, Netherlands. [2]Faculty of Arts, Science and Technology, University of Northampton, Northampton NN1 5PH, UK. Email: c.j.van.der.kooi@rug.nl

Published by AAAS

Evolution of angiosperms according to molecular and fossil evidence

Fossil and molecular evidence lead to conflicting conclusions about the timing of the origin of flowering plants. Fossil evidence suggests that flowering plants arose near the beginning of the Cretaceous, but molecular analyses date the origin much earlier, in the Triassic.



the scheme of angiosperm evolution, because even if they occurred at this earlier period, angiosperms were not a dominant plant group in the Jurassic. By contrast, Bennettitales and other early seed plants were ecologically dominant in Late Triassic to Jurassic floras, indicating that the transition to insect pollination in angiosperms arose through these gymnosperm groups. These possibilities are more complex than the standard scenarios that envisioned a progression from primitive wind pollination to advanced insect pollination. They hint at a richer ecological milieu of more complex interactions between species than had previously been appreciated, including insect groups that are currently much less important as pollinators, such as scorpionflies (Mecoptera) (*11*).

The timing of flowering-plant origins also provides a minimum age for the evolution of their most prominent feature: flowers. Insect pollination in many extant gymnosperms (e.g., cycads, *Ephedra*, *Gnetum*) is facilitated mainly by scent rather than by visual attraction. The same may have been true of the extinct gymnosperms, but because scent does not fossilize, it may be impossible to ever know. However, if the reproductive structures of these extinct gymnosperms functioned in a manner similar to their living relatives, with odor predominating, then the increasing importance of visual-based cues to attract pollinators in angiosperms could be one of the defining features of angiosperm evolution and success. Further, if floral structures predate some speciose orders of flower-visiting insects, perhaps flower features have shaped trait evolution in these large insect groups.

There are clear examples of coevolution of specific floral and pollinator morphological characteristics in some systems, such as floral tube length and pollinator tongue length (*12*). What about floral features such as color and scent? For example, perhaps floral color and scent evolved to match pollinator vision and olfaction, or vice versa. Alternatively, signal production may have evolved synchronously with detection. The basic principles of color vision in insects, such as the possession of three types of photoreceptors (ultraviolet, blue, green), seem to predate flowers regardless of whether they arose during the Triassic or later (*13*). Because color vision is also used for key behaviors such as detecting potential mates and predators and finding oviposition (egg-laying) sites, the evolution of color vision is unlikely to be driven by flower colors.

A similar ancestral origin of olfaction compared to scent production was documented in a group of plants pollinated by scarab beetles (*14*), where odor reception by pollinators predates production of the scent signal by plants. However, behavioral aspects of olfaction or color vision, such as innate color preferences that shape foraging behavior in various insect groups (*15*), may have evolved later, in response to floral signals. All of this depends on the timing of the evolution of flowering plants as well as the order of evolutionary events that led to insect pollination. If insect-pollinated gymnosperms predate angiosperms, for example, then it may be possible to trace the origin of these visual and olfactory traits to long-extinct clades of plants that once dominated terrestrial floras.

Future paleontological discoveries will undoubtedly reveal additional fossils, and the use of complementary sequencing approaches and more sophisticated evolutionary models will help to mitigate the limitations imposed by the rampant polyploidy in plants that frequently hinders analysis of nuclear genes. Whether Darwin's question about the timing of flowering-plant evolution and radiation will ever be answered remains a mystery, but clearly this question and its ecological implications for understanding insect pollination are complicated. ■

REFERENCES AND NOTES

1. H.-T. Li *et al.*, *Nat. Plants* **5**, 461 (2019).
2. M. Coiro, J. A. Doyle, J. Hilton, *New Phytol.* **223**, 83 (2019).
3. S. Magallón, S. Gómez-Acevedo, L. L. Sánchez-Reyes, T. Hernández-Hernández, *New Phytol.* **207**, 437 (2015).
4. P. S. Herendeen, E. M. Friis, K. R. Pedersen, P. R. Crane, *Nat. Plants* **3**, 17015 (2017).
5. J. M. Beaulieu, B. C. O'Meara, P. Crane, M. J. Donoghue, *Syst. Biol.* **64**, 869 (2015).
6. J. Barba-Montoya *et al.*, *New Phytol.* **218**, 819 (2018).
7. S. Cardinal, B. N. Danforth, *Proc. R. Soc. B* **280**, 20122686 (2013).
8. T. J. B. van Eldijk *et al.*, *Sci. Adv.* **4**, e1701568 (2018).
9. A. Y. Kawahara *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22657 (2019).
10. J. Ollerton, *Annu. Rev. Ecol. Evol. Syst.* **48**, 353 (2017).
11. D. Ren *et al.*, *Science* **326**, 840 (2009).
12. B. Anderson, S. D. Johnson, *Evolution* **62**, 220 (2008).
13. L. Chittka, *Isr. J. Plant Sci.* **45**, 115 (1997).
14. F. P. Schiestl, S. Dötterl, *Evolution* **66**, 2042 (2012).
15. C. J. van der Kooi, A. G. Dyer, P. G. Kevan, K. Lunau, *Ann. Bot.* **123**, 263 (2019).

ACKNOWLEDGMENTS

C.J.v.d.K. is funded by a Veni grant from the Dutch NWO (016.Veni.181.025) and AFOSR/EOARD (FA9550-15-1-0068). We thank E. Zinkstok for help with the figure.

10.1126/science.aay3662

GRAPHIC: ADAPTED BY N. CARY/*SCIENCE* FROM E. ZINKSTOK

Published by AAAS

PHOTOVOLTAICS

Perovskite solar cells take a step forward

A new encapsulation technique helps move a photovoltaic toward commercialization

By **Emilio J. Juarez-Perez**[1,2,3] and **Marta Haro**[3]

Today's monocrystalline silicon solar cells have their throne on the roofs of our houses. In the past decade, however, perovskite solar cells (PSCs) show impressive advances with a high power conversion efficiency (PCE) of 25.2% (*1*) and low fabrication cost, which make this technology promising for further advances in decarbonization energy models (*2*). Yet the life cycle of PSCs needs to be increased for market integration. Poor stability is the main impediment to commercializing this technology. Thus, great effort has been focused on the causes and mechanisms of degradation, many of which can be mitigated or minimized with encapsulation. Various strategies have been proposed to increase PSCs' operational stability, which is affected by moisture, oxidation, heat, light, and other factors (*3*, *4*). On page 1328 of this issue, Shi *et al.* (*5*) report a successful encapsulation procedure for hybrid PSCs.

"...vapors retained by encapsulation could regenerate perovskite during night hours..."

The standard test to assess the long-term operation of photovoltaic terrestrial flat-plate modules in open-air climates is referred to as IEC 61215:2016. The IEC 61215 standard was developed by the International Electrotechnical Commission (IEC) for working photovoltaic devices that incorporate robust inorganic light-harvesting materials such as silicon, cadmium telluride, copper indium selenide, and copper indium gallium selenide. The stability test must consider all components and encapsulating materials as a whole. Achieving a high score on a solar cell stress test has been challenging for organic-inorganic hybrid perovskite-based devices (*6*). The device developed by Shi *et al.* retained >95% of its initial efficiency after undergoing the harsh environmental conditions of the Damp Heat and Humidity Freeze tests. By doing so, their device exceeded the requirements of IEC 61215:2016.

The key strategy for the authors' success was the polymeric "blanket-cover" encapsulation that simultaneously prevented moisture/air ingress into the device and suppressed volatile organic gas release from the hybrid perovskite. The analysis of outgassing from several perovskite and precursor assemblies, monitored by coupled gas chromatography–mass spectrometry, was essential for determining and implementing the best encapsulation strategy. Beyond the simplicity of this encapsulation approach, the authors demonstrated the feasibility of developing low-cost packaging techniques for PSCs, in contrast with more technically challenging and expensive processes of forming a protective barrier, such as atomic layer deposition or chemical vapor deposition.

A reproducible high PCE (>20%) combined with the authors' demonstration that their PSC device can pass the harsh IEC 61215:2016 tests might suggest that PSCs are finally ready for commercialization. Unfortunately, each of these achievements, high efficiency and high stability, still has a downside. Deploying PCEs well above 20% is still problematic when scaling up the active area for these photovoltaics. This apparently is specific to PSCs, as it is practically nonexistent with wafer or other thin film–based photovoltaics. Moreover, the authors observed an enigmatic threshold temperature during the regular working condition for a solar cell under constant illumination, the so-called maximum power point test (MPPT). Once this temperature is exceeded, a prominent drop in efficiency initiates and drives the irreversible loss of the device. Shi *et al.* estimated this temperature threshold to be ~35° to 40°C despite their devices passing the IEC tests, which may require further specific thermal stability studies on PSCs.

This temperature threshold has been noted in a previous MPPT that successfully tested PSCs for 4000 hours of continuous illumination (*7*). Surprisingly, the threshold temperature for the earlier test was also ~40°C, even though nonmixed or pure methylammonium (MA) cation-type perovskite was used. This is in contrast to the mixed FAMA and CsFAMA perovskites [containing MA, FA (formamidinium), and cesium] used by Shi *et al.* A much higher temperature threshold would be expected for FAMA and CsFAMA because of their higher FA content. This analogous temperature threshold could be due to degradation factors such as MA content, diiodine sublimation, or disruptive ion migration. Determining this threshold in MA-free perovskites could provide insight, along with determining whether full inorganic halide perovskite-based devices are free of this threshold at MPPT conditions under 1-sun illumination. Standard stability tests are designed to stress all plausible weak points of the device and offer useful data for interlaboratory comparisons. Although considerable efforts to design specific stability stress tests for hybrid perovskites are under way (*8*, *9*), there is a lack of standardization in testing procedures, including specific active control of temperature during MPPT to find the fatal threshold temperature of the device.

Shi *et al.* provide further insight into the challenge of long-term stabilization for PSC devices by reporting the outgassing vapors generated above the ~35° to 40°C temperature threshold. The outgassing vapors retained by encapsulation could regenerate perovskite during night hours when the device cools down, but this is not possible if these vapors react irreversibly with other components of the device, such as selective contacts and metal electrodes. The chemistry of halide perovskite is quite different from the classic robust inorganic crystalline materials prepared at high temperature and used in commercial photovoltaics. Perovskite is an organic-inorganic hybrid semiconductor material, and this hybrid character can be assumed to be the clue to developing long-lasting PSCs for commercialization. ■

[1]Aragonese Foundation for Research and Development (ARAID), Government of Aragon, Zaragoza 50018, Spain. [2]Institute of Nanoscience of Aragon, University of Zaragoza, Zaragoza 50018, Spain. [3]Institute of Materials Science of Aragon (ICMA), University of Zaragoza, Zaragoza 50009, Spain. Email: ejjuarezperez@unizar.es

REFERENCES AND NOTES

1. National Renewable Energy Laboratory, Best Research-Cell Efficiency Chart (2020); www.nrel.gov/pv/cell-efficiency.html.
2. European Perovskite Initiative, White Paper (2020); https://epki.eu/.
3. S. He *et al.*, *Sci. Eng. Rep.* **140**, 100545 (2020).
4. J. Bisquert, E. J. Juarez-Perez, *J. Phys. Chem. Lett.* **10**, 5889 (2019).
5. L. Shi *et al.*, *Science* **368**, eaba2412 (2020).
6. R. Cheacharoen *et al.*, *Sustain. Energy Fuels* **2**, 2398 (2018).
7. M. B. Islam, M. Yanagida, Y. Shirai, Y. Nabetani, K. Miyano, *Sol. Energy Mater. Sol. Cells* **195**, 323 (2019).
8. M. V. Khenkin *et al.*, *Nat. Energy* **5**, 35 (2020).
9. A. García-Fernández *et al.*, *Small Methods* **2**, 1800242 (2018).

ACKNOWLEDGMENTS

M.H. is supported by the MICIU (Spain) through a Ramon y Cajal Fellowship (grant RYC-2018-025222-I).

10.1126/science.abc5401

Published by AAAS

INFECTIOUS DISEASES

Dating the emergence of human pathogens

Ancient genomes can narrow the search for the sources of zoonotic transmissions

By **Simon Y. W. Ho**[1] and **Sebastián Duchêne**[2]

Understanding the emergence and evolution of human pathogens plays a pivotal role in epidemiology and in predicting the trajectories of outbreaks. The application of phylogenetic methods to pathogen genomes has provided a range of insights into their evolutionary dynamics (*1*). In many cases, phylogenetic methods can use the sampling dates of the genomes to reconstruct the evolutionary time scales of viruses, bacteria, and other pathogens. Ancient genomes can increase the power of these approaches by narrowing the estimated time window of pathogen emergence and by augmenting the evolutionary temporal signal in the genetic data. On page 1367 of this issue, Düx *et al.* (*2*) show how a century-old genome of *Measles morbillivirus*, extracted from human lung tissue, can help efforts to pinpoint the time of emergence of measles.

A distinctive feature shared by many human pathogens is that their evolution can be observed over epidemiological time scales. These pathogens can undergo measurable genetic change even between successive sampling dates. The evolutionary rates of these pathogens can be estimated by analyzing their genomes while taking the sampling dates into account (*3*). Datasets that are amenable to such analysis are referred to as having "temporal signal" (*4*).

Strong temporal signals are often detected in analyses of rapidly evolving RNA viruses. However, they can also be detected in genomic data from more slowly evolving bacteria, as demonstrated in studies of the tuberculosis pathogen *Mycobacterium tuberculosis* (*5*) and *Salmonella enterica* (*6*). Conversely, temporal signals have proven to be rather more elusive in some pathogens, such as the bacterium that causes leprosy, *Mycobacterium leprae* (*7*).

The temporal signal can be amplified by widening the sampling period, which is most effectively done by sequencing genomes from samples in historical collections or from archaeological remains. These efforts have been substantially aided by advances in molecular techniques and sequencing technology. However, obtaining genomic data from RNA viruses (such as the measles virus), which degrade rapidly in the environment, continues to be extremely challenging (*8*). In this regard, the sequencing of a century-old genome of the measles virus by Düx *et al.* is a profound achievement.

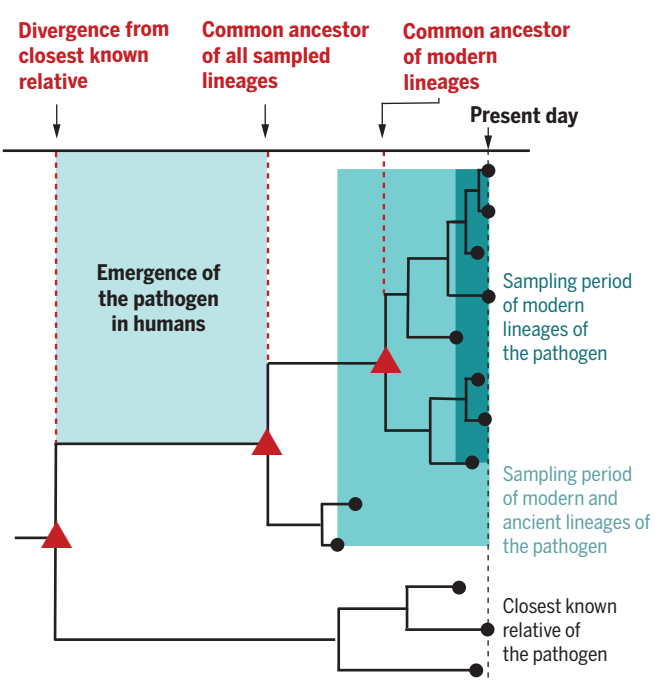
Increasing the temporal signal in the measles virus genomic dataset leads to a more accurate estimate of the evolutionary rate of the pathogen, which allows inference of the date of the common ancestor of the sampled lineages with greater confidence. This date places a minimum bound on the emergence of the pathogen in humans (see the figure), because the common ancestor implies a human-to-human transmission event.

The inclusion of ancient genomes can also push back the date of the common ancestor of the pathogen by sampling extinct lineages (*8*). In this way, ancient genomes are able to cast light on a hidden period of evolutionary history, an effect that is seen in the analysis of the measles virus by Düx *et al.* and in previous studies of the bacterium *S. enterica* (*6*), smallpox (variola) virus (*9*), and the plague bacterium *Yersinia pestis* (*10*). Sampling extinct lineages is particularly important for pathogens that have undergone recent reductions in genetic diversity. These include viruses that have been subject to large-scale vaccination programs, such as measles (as reported by Düx *et al.*), as well as those that have been targeted by extensive treatment and prevention strategies, such as human immunodeficiency virus type 1 (HIV-1) (*11*). In these cases, there is likely to have been a substantial assortment of extinct lineages that have not left any traces in the present-day population of the pathogen.

The maximum age of a pathogen's emergence in humans is bounded by its evolutionary divergence from its closest known relative. Düx *et al.* estimated that the measles virus diverged from its closest relative, the cattle pathogen *Rinderpest morbillivirus*, more than two millennia ago. The authors then posited that the establishment of measles in humans coincided with population growth in Eurasia between 2000 and 2500 years ago, although it could have occurred much more recently. This estimate is consistent with other studies that have suggested that the rise of human settlements facilitated the spread of some infectious pathogens, including the plague bacterium (*12*).

Pinpointing the emergence of human pathogens

The emergence of human pathogens can be bracketed between their divergence from their closest relatives and the common ancestor of all sampled lineages of the pathogen. Including ancient genomes expands the sampling period and can reveal extinct lineages of the pathogen.



GRAPHIC: H. BISHOP/*SCIENCE*

[1]School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia. [2]Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, Australia.
Email: simon.ho@sydney.edu.au

Published by AAAS

For pathogens with sources that are less well understood, the window of uncertainty can be narrowed by searching for close relatives of the pathogen. Such a search can be targeted at candidate reservoir hosts or can involve a broader survey of pathogen diversity across wildlife. For example, the growing application of metagenomics (nontargeted genome sequencing) approaches has led to a staggering expansion of the known diversity of viruses (*13*), while revealing that animal hosts can carry a wide range of viruses in abundance. Nevertheless, huge swathes of the virosphere remain unexplored.

Phylogenetic dating analyses of pathogens have come under the spotlight in the ongoing coronavirus disease 2019 (COVID-19) pandemic. The closest known relative of the causative agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a coronavirus from a horseshoe bat, but these two viruses diverged from each other several decades ago (*14*). The most recent common ancestor of sampled SARS-CoV-2 genomes has been dated to late November to early December 2019 (*15*). The gap between these two events has hindered attempts to identify the reservoir host of SARS-CoV-2 and the timing of the zoonotic spillover into human hosts.

Further genomic data from historical and ancient samples, along with more comprehensive and intensive surveys of viruses harbored in wildlife, will lead to continued refinements of the time scales of emergence and evolution of human pathogens. In turn, these refinements will improve our understanding of the circumstances under which pathogens emerge in their hosts and the mechanisms by which they do so. ■

REFERENCES AND NOTES

1. O. G. Pybus, A. Rambaut, *Nat. Rev. Genet.* **10**, 540 (2009).
2. A. Düx *et al.*, *Science* **368**, 1367 (2020).
3. A. Rieux, F. Balloux, *Mol. Ecol.* **25**, 1911 (2016).
4. R. Biek, O. G. Pybus, J. O. Lloyd-Smith, X. Didelot, *Trends Ecol. Evol.* **30**, 306 (2015).
5. F. Menardo, S. Duchêne, D. Brites, S. Gagneux, *PLOS Pathog.* **15**, e1008067 (2019).
6. Z. Zhou *et al.*, *Curr. Biol.* **28**, 2420 (2018).
7. S. Duchêne *et al.*, *Microb. Genom.* **2**, e000094 (2016).
8. M. A. Spyrou, K. I. Bos, A. Herbig, J. Krause, *Nat. Rev. Genet.* **20**, 323 (2019).
9. A. T. Duggan *et al.*, *Curr. Biol.* **26**, 3407 (2016).
10. N. Rascovan *et al.*, *Cell* **176**, 295 (2019).
11. S. Gryseels *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12222 (2020).
12. M. A. Spyrou *et al.*, *Nat. Commun.* **9**, 2234 (2018).
13. Y.-Z. Zhang, M. Shi, E. C. Holmes, *Cell* **172**, 1168 (2018).
14. R. Lu *et al.*, *Lancet* **395**, 565 (2020).
15. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, *Nat. Med.* **26**, 450 (2020).

ACKNOWLEDGMENTS

S.Y.W.H. and S.D. were funded by the Australian Research Council.

10.1126/science.abc5746

NEUROSCIENCE

Getting a grip on touch receptors

Meissner corpuscles are anatomically complex mechanosensors critical for tactile acuity

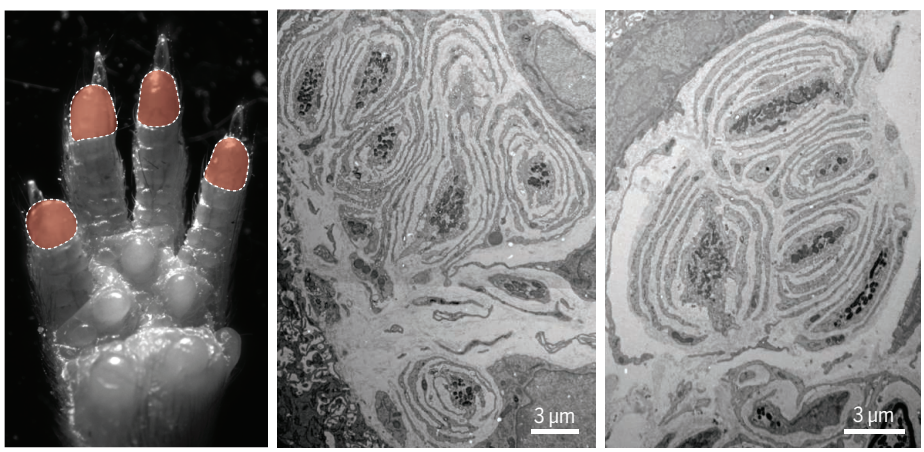
By **Kara Marshall** *and* **Ardem Patapoutian**

Skin sets the boundaries of the body, and neuron endings within the skin translate pressure into percept. When exploring a surface with our hands, exquisitely mechanosensitive touch neurons can detect tiny wrinkles in texture down to 10 nm (*1*) or mere 3° differences in edge orientation (*2*). These feats are accomplished by diverse sensory cells that are tuned to capture the subtleties of our external world. On page 1330 of this issue, Neubarth *et al.* (*3*) characterize Meissner corpuscles, mechanosensory end organs in the skin that detect touch. The authors report that in mice, two neuron types comprise a single Meissner corpuscle and that differences in their stimulus thresholds, electrical conduction velocity, and response patterns could underly the sensitivity range of Meissner corpuscles. They also demonstrate that these receptors are critical for tactile acuity.

The skin is vast. Touch receptors cover every bodily surface, and their varied morphologies point to equally diverse functions. Merkel cell–neurite complexes that detect gentle touch cup epidermal ridges in the fingertips but elsewhere terminate closer to the skin surface. Lanceolate endings encircle the base of each hair and detect hair movement. Pacinian corpuscles sit deep in the skin whereas Meissner corpuscles are superficial, but both detect vibrations (*4*). This distributed array, and the lack of specific markers, has made it difficult to study individual touch receptor types.

Meissner neuron afferents are rapidly adapting, meaning that they have brief responses to sustained pressure applied to the skin. This response profile is ideal for detecting movement and small slips that occur when scanning an object (*5*). Thus, Meissner corpuscles have been hypothesized to be critical for tactile acuity, but the only way to confirm this is to selectively remove them. Neubarth *et al.* accomplish this in the mouse. The breadth of their findings was enabled by a variety of tools, ranging from electron microscopy to naturalistic behavioral paradigms.

The morphology of neuron endings in the skin is inherently related to function. For example, hair follicles act as levers for the neurons that surround them, and ample lamellar wrappings from Schwann cells create a filter that shields Pacinian corpuscles from low-frequency vibrations. Meissner corpuscles are also wrapped in lamellae and poised to detect superficial stimuli, as they sit at the top of dermal papillae. Genetic tools to distinguish distinct touch receptor populations have not been available until recently (*6, 7*). By combining known Meissner neuron marker genes, Neubarth



Mouse digits have a high density of Meissner corpuscles (left). Electron microscopy of Meissner corpuscles reveals that TrkB-expressing neurons (middle) have more Schwann cell wrappings than Ret-expressing neurons (right).

PHOTOS: NEUBARTH *ET AL.* (3)

Howard Hughes Medical Institute, Department of Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA 92037, USA.
Email: ardem@scripps.edu; marshall@scripps.edu

Published by AAAS

et al. found that two genetically distinct neurons intertwine in a single Meissner corpuscle. Electron microscopy revealed that neuron endings that express Ret (rearranged during transfection) receptors had fewer lamellar wrappings compared to neuron endings that express TrkB (tropomyosin receptor kinase B) receptors (see the figure).

Lamellar wrappings dampen vibration responses of Pacinian corpuscles, and removing lamellae converts rapidly adapting responses into sustained ones (*8*). Indeed, in vivo recordings revealed functional differences between Meissner afferent types: Compared to TrkB-expressing neurons, Ret-expressing neurons had higher thresholds, faster electrical conduction velocity, and sometimes sustained responses, but lacked "off" responses to stimulus removal. It is not clear whether these properties are conferred by the modest variations in lamellar structure, differences in protein expression, or a combination of mechanisms. The findings do suggest that a Meissner corpuscle is two sensors joined into one, which offers an intriguing mechanism by which a single receptor can achieve greater dynamic range.

What information do Meissner corpuscles convey? For years, the standard tool for studying mouse touch has been von Frey filaments, which are a set of probes with differing stiffness. The readout for whether a mouse feels these probes is paw movement when poked, but mice vary in their motivation, and unexpected pokes do not represent normal tactile experiences. Scientists have recently developed newer assays to probe touch in more creative ways. For example, mice prefer still surfaces to vibrating ones and will ceaselessly work to remove tape stuck on their back (*9*). Tasks that probe recognition of unfamiliar objects can be used for testing texture detection, and air puffs can test hair movement sensation (*10*). These developments have expanded our understanding of touch, but they fall short of psychometric tests for skill. Moreover, they offer a limited understanding of how mice use touch in natural settings.

Neubarth *et al.* designed an operant conditioning paradigm using a water reward delivered after a gentle touch, which allowed them to build psychometric curves. Mice lacking Meissner corpuscles were far less skilled at detecting gentle indentation in this task. Additionally, behavioral analysis of mice handling sunflower seeds revealed that Meissner corpuscles are critical for fine sensorimotor control. This experimental evidence connects a specific touch receptor with naturalistic tactile behavior. The deficits observed in these animals, despite having many other functional touch receptors, emphasize that Meissner corpuscles are indispensable for creating the full spectrum of touch sensation. Future work could disentangle the differential contributions of each Meissner neuron subtype to these tasks.

Rapidly adapting responses from Meissner corpuscles are ideal for conveying dynamic touch, and their sensitivity to movement and slip have long made them candidates for fine-tuning grip. Primate species that spend more time handling fruit have higher densities of Meissner corpuscles in their palm skin (*11*). The findings of Neubarth *et al.* solidify this role. Nonetheless, these receptors are insensitive to static stimuli and poorly resolve spatial details (*4*). Many studies have confirmed that spatial discrimination, particularly of coarse or static stimuli, is better served by slowly adapting responses from Merkel cell–neurite complexes (*4*). Behavioral tasks that select for spatial discrimination will be crucial to fully understand the functional contributions of different touch receptors.

These data raise several questions. Where does the mechanotransduction ion channel PIEZO2 (Piezo type mechanosensitive ion channel component 2), which is necessary for touch sensation (*9, 12*), sit in this structure? Neubarth *et al.* and others (*13*) identified small lamellar openings that expose the naked axon to the extracellular matrix. Might these openings be sites of mechanotransduction? It is also unclear what benefit dual innervation offers. Computational modeling suggests that the organization observed for Meissner corpuscles increases tactile information and acuity (*3*). This could be a broader theme among touch receptors. Different neurons are neatly interdigitated in lanceolate endings (*13*), and Merkel cells occasionally have atypical neurons wandering nearby (*14*). These neurons are not as intimately intertwined as in Meissner corpuscles, but perhaps they function similarly. These exciting findings pave the way for understanding the full complexity of touch receptor function. ■

REFERENCES AND NOTES

1. L. Skedung *et al.*, *Sci. Rep.* **3**, 2617 (2013).
2. J. A. Pruszynski, J. R. Flanagan, R. S. Johansson, *eLife* **7**, e31200 (2018).
3. N. L. Neubarth *et al.*, *Science* **368**, eabb2751 (2020).
4. V. E. Abraira, D. D. Ginty, *Neuron* **79**, 618 (2013).
5. K. O. Johnson, *Curr. Opin. Neurobiol.* **11**, 455 (2001).
6. S. Bourane *et al.*, *Neuron* **64**, 857 (2009).
7. W. Luo, H. Enomoto, F. L. Rice, J. Milbrandt, D. D. Ginty, *Neuron* **64**, 841 (2009).
8. W. R. Loewenstein, M. Mendelson, *J. Physiol.* **177**, 377 (1965).
9. S. S. Ranade *et al.*, *Nature* **516**, 121 (2014).
10. L. L. Orefice *et al.*, *Cell* **166**, 299 (2016).
11. J. N. Hoffmann, A. G. Montag, N. J. Dominy, *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* **281**, 1138 (2004).
12. A. T. Chesler *et al.*, *N. Engl. J. Med.* **375**, 1355 (2016).
13. L. Li, D. D. Ginty, *eLife* **3**, e01901 (2014).
14. C. M. Reinisch, E. Tschachler, *Ann. Neurol.* **58**, 88 (2005).

ACKNOWLEDGMENTS

We thank R. Hill for helpful editing.

10.1126/science.abc7610

ORGANIC CHEMISTRY

Closing the radical gap in chemical synthesis

Unstable radical intermediates are harnessed in a microfluidic electrochemical cell

By **Jian-Quan Liu[1,2], Andrey Shatskiy[2], Markus D. Kärkäs[2]**

Chemists now face the daunting task of identifying synthetic methods that are more practical, scalable, and sustainable. Approaches include the development of safer, more readily available reactants and reagents as well as discoveries in method development that provide more effective and selective synthetic transformations (*1, 2*). Conventionally, organic synthesis has been conducted through batch processes, but the use of continuous-flow technologies can allow straightforward scale-up, facilitate operational simplicity, and reduce occupational hazards and industrial waste streams (*3*). On page 1352 of this issue, Mo *et al.* (*4*) demonstrate that microfluidic electrochemical technologies can be applied to single-electron transfer (SET) redox-neutral reactions. Their strategy enables the efficient formation and utilization of free-radical intermediates by placing the anode and cathode in close proximity to overcome radical instability. These versatile intermediates were leveraged in Minisci-type reactions of electron-deficient heterocyclic compounds, as well as radical-radical and nickel-catalyzed sp^2 carbon–oxygen cross-couplings.

Students of organic chemistry are mainly taught the idea of assembling organic compounds by reacting a pair of oppositely charged species, such as electrophiles with nucleophiles (*5*). However, these ionic transformations often have considerable activation barriers that must be overcome

[1]School of Chemistry and Materials Science, Jiangsu Normal University, Xuzhou, Jiangsu 221116, China. [2]Department of Chemistry, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. Email: karkas@kth.se

Published by AAAS

thermally. Alternatively, lower-energy pathways can be accessed through the use of catalysis or additives. By contrast, many emerging approaches rely on free-radical chemistry. In organic synthesis, radical reactions historically have been perceived as uncontrollable and impractical, but this perception has gradually shifted, mainly because of the conceptual advances in the areas of photoredox catalysis (*6, 7*) and electrosynthesis (*8, 9*).

Redox-modulating techniques allow chemists to target functional groups in a molecule on the basis of their different redox potentials, thereby enabling the mild and selective generation of radicals. Compared to conventional two-electron chemistry, using radicals opens up alternative reactivity and selectivity profiles and can greatly simplify synthesis through non-traditional bond construction. The revitalized interest in radicals has propelled their incorporation in contemporary chemical synthesis (*10, 11*).

Reactions that proceed through radical intermediates can be classified as net oxidations, net reductions, or redox neutral. In the first two cases, stoichiometric reagents must be used as electron donors or acceptors. In redox-neutral reactions, the balance of electron-transfer steps between the reacting molecular partners allows construction of complex molecules in a more atom-economical fashion (*12*). In one of the redox-neutral approaches, both reactive partners are transformed into free-radical intermediates, and bond formation proceeds through radical-radical coupling.

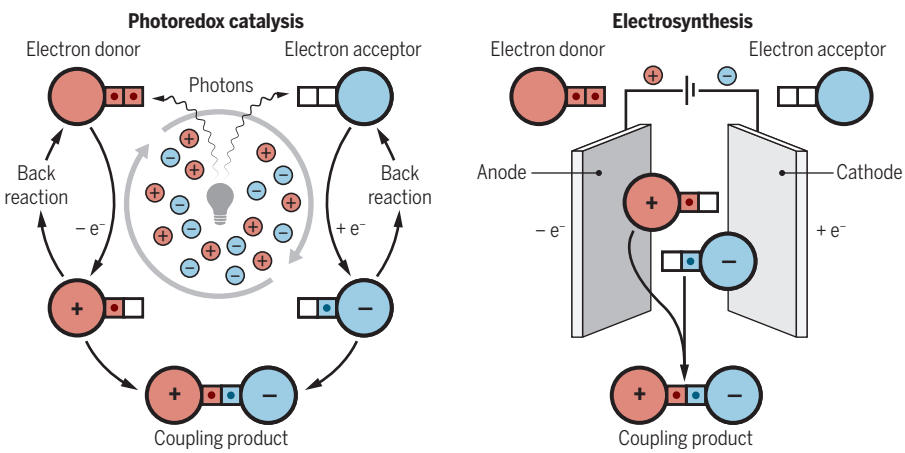
The activation barrier in such reactions predominantly stems from the oxidation and reduction steps and can be surmounted with the aid of photoredox catalysis or electrosynthesis. Either the energy of photons in photoredox setups or the potential energy between the electrodes in an electrosynthetic cell can drive the formation of radical intermediates. An important distinction between the two approaches is the spatial separation between the oxidation and reduction reaction sites (see the figure). In photoredox catalysis, the molecular photocatalyst acts as both the oxidizing and reducing agent and is freely diffusing in solution. In an electrosynthetic cell, the oxidation and reduction reactions are confined to the surface of the cell electrodes.

This distinction leads to profound differences in reactivity patterns and intrinsic limitations observed in both systems. For example, back-electron transfer can compete with the productive reaction pathways in photocatalytic systems, and the short half-life of certain free-radical species can restrict product formation in electrosynthetic cells. The latter challenge has been successfully addressed in the work by Mo *et al.* by introduction of a microfluidic redox-neutral electrochemistry (μRN-eChem) platform inspired by the previous developments in continuous-flow electrosynthesis (*13*, *14*). The disclosed approach relies on an extremely small interelectrode gap (down to 25 μm) in the electrosynthetic cell that substantially decreases the diffusion time between electrodes and intensifies the mass transport of unstable free-radical coupling partners generated on the electrodes. The decreased interelectrode gap also eliminates the need for a supporting electrolyte and redox mediators, commonly used to facilitate charge transfer between the electrodes and the species in solution.

These improvements allowed Mo *et al.* to use the μRN-eChem platform with reactions in which the free-radical intermediates are too unstable to allow their efficient coupling in conventional electrosynthetic cells or in photoredox systems. They executed redox-neutral coupling reactions with unstable carbon- and nitrogen-based radical intermediates, illustrating that the immense potential of radical-chemistry manifolds can be harnessed with an appropriate electrosynthetic cell design. To date, complex radical-based transformations were mainly demonstrated in photoredox settings, and far fewer synthetically challenging reactions were executed with electrosynthesis (*15*). The current work strives to eliminate this difference and provides synthetic chemists with a powerful tool for expanding the scope of redox-neutral free-radical reaction manifolds in electrochemical settings. ■

Approaches to redox-neutral radical-radical coupling

Photoredox and electrosynthetic approaches to radical-radical coupling reactions differ in fundamental ways. Both generate radicals through single-electron transfers that can undergo otherwise difficult coupling reactions.



Photocatalytic systems

The photogenerated oxidized and reduced radical species are distributed in solution and are subject to back reaction to the neutral species.

Electrosynthetic systems

The oxidation and reduction reactions are confined to the electrode surfaces. In the microfluidic system of Mo *et al.*, these species come together rapidly for reaction.

GRAPHIC: C. BICKEL/*SCIENCE*

REFERENCES AND NOTES

1. A. M. Armaly, Y. C. DePorre, E. J. Groso, P. S. Riehl, C. S. Schindler, *Chem. Rev.* **115**, 9232 (2015).
2. L. A. Morrill, R. B. Susick, J. V. Chari, N. K. Garg, *J. Am. Chem. Soc.* **141**, 12423 (2019).
3. M. B. Plutschack, B. Pieber, K. Gilmore, P. H. Seeberger, *Chem. Rev.* **117**, 11796 (2017).
4. Y. Mo *et al.*, *Science* **368**, 1352 (2020).
5. E. J. Corey, X.-M. Cheng, *The Logic of Chemical Synthesis* (Wiley, 1995).
6. M. H. Shaw, J. Twilton, D. W. C. MacMillan, *J. Org. Chem.* **81**, 6898 (2016).
7. M. D. Kärkäs, J. A. Porco Jr., C. R. J. Stephenson, *Chem. Rev.* **116**, 9683 (2016).
8. M. Yan, Y. Kawamata, P. S. Baran, *Chem. Rev.* **117**, 13230 (2017).
9. M. D. Kärkäs, *Chem. Soc. Rev.* **47**, 5786 (2018).
10. J. M. Smith, S. J. Harwood, P. S. Baran, *Acc. Chem. Res.* **51**, 1807 (2018).
11. K. J. Romero, M. S. Galliher, D. A. Pratt, C. R. J. Stephenson, *Chem. Soc. Rev.* **47**, 7851 (2018).
12. J. Xie, H. Jin, A. S. K. Hashmi, *Chem. Soc. Rev.* **46**, 5193 (2017).
13. D. Pletcher, R. A. Green, R. C. D. Brown, *Chem. Rev.* **118**, 4573 (2018).
14. M. Atobe, H. Tateno, Y. Matsumura, *Chem. Rev.* **118**, 4541 (2018).
15. A. Shatskiy, H. Lundberg, M. D. Kärkäs, *ChemElectroChem* **6**, 4067 (2019).

ACKNOWLEDGMENTS

Financial support from KTH Royal Institute of Technology to M.D.K. is gratefully acknowledged. The Wenner-Gren Foundations and the Olle Engkvist Foundation are kindly acknowledged for postdoctoral fellowships to J.-Q.L. and A.S., respectively. These authors contributed equally to this work.

10.1126/science.abc2985

Published by AAAS

MEDICINE

Drug modulation by nuclear condensates

Concentration of antineoplastic agents into spatial compartments influences activity

By **Aaron D. Viny**[1,2] *and* **Ross L. Levine**[1,2]

Many mysteries remain about the eukaryotic nucleus. In the past decade, much has been discovered anew about the three-dimensional (3D) organization of the nucleus and the dynamic interactions therein that influence cellular function. Studies have uncovered the critical roles that topologic structure, histone modifications, and DNA modifications play in regulating transcription. By contrast, the understanding of interactions among proteins, RNA, and chromatin in macromolecular assemblies is less developed. These condensates are the molecular basis for discrete nuclear spatial organization of active and repressive chromatin as well as distinct nuclear structures such as the nucleolus (*1, 2*). On page 1386 of this issue, Klein *et al.* (*3*) begin to dissect the functional relevance of nuclear condensates in regulating the distribution and activity of fluorescent-labeled antineoplastic agents with specific nuclear localization profiles. This finding could have wide-ranging implications for the understanding of pharmacologic mechanisms and has substantive consequences for therapeutic development, drug delivery, and target engagement.

Recent technologies have allowed for appreciation of hierarchical 3D structures within the nucleus that have both known and unknown functional importance. Most widely accepted is the principle of spatial separation of chromatin into regions of active transcription (euchromatin). These regions share the property of deoxyribonuclease (DNase) or transposase accessibility and the presence of transcriptional coactivators, including the mediator (MED) complex. These genomic regions are spatially separate from regions of transcriptional inactivity (heterochromatin), densely packed histone-bound chromatin, often demarcated by repressive histone methylation marks.

Beyond euchromatin and heterochromatin, other segregated nuclear subcompartments have been described as self-organizing condensates. Through careful assessment of nuclear condensates (*4, 5*), Klein *et al.* observed that fluorescent-labeled small molecules had distinct and specific spatial aggregation in different nuclear compartments. Fluorescent-labeled cisplatin, a chemotherapy drug that damages and cross-links DNA, was observed to aggregate in mediator of RNA polymerase II transcription subunit 1 (MED1) transcriptional condensates while diffusing freely through heterochromatin protein 1 (HP1)–demarcated heterochromatin. Upon locus-specific assessment of cisplatin-induced DNA damage (platination) across the genome, only the DNA contained in MED1 condensates had high amounts of cisplatin incorporation. This finding suggests specificity of the pharmacologic effect of cisplatin to a distinct type of condensate. Additionally, the differential effect on platinum incorporation was lost with bromodomain-containing protein 4 (BRD4) inhibition, which disrupts MED1 condensates.

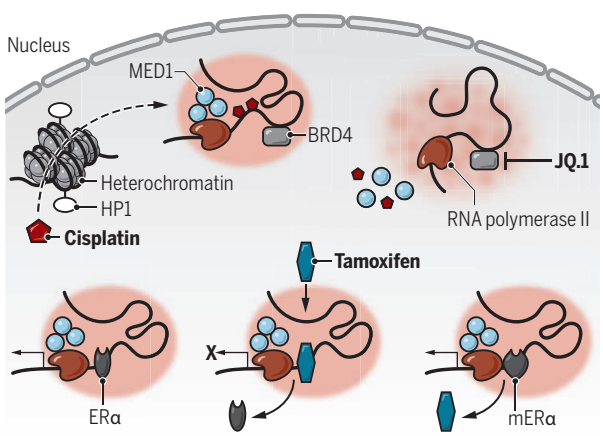
The authors also assessed the localization of tamoxifen and found that this targeted endocrine chemotherapy aggregated in MED1 condensates, competing with and evicting its target, estrogen receptor α (ERα), from these intranuclear structures. In human breast cancer cells, they found that ERα interacts with the *MYC* locus in MED1 condensates; this association was abrogated by tamoxifen and condensate accumulation. Notably, patient-derived ERα mutations that induce tamoxifen resistance led to their retention in MED1 condensates, rendering tamoxifen unable to outcompete oncogenic ERα.

Klein *et al.* demonstrate that a subset of small-molecule therapeutics can have differential enrichment in nuclear compartments (see the figure). Moreover, functional studies have demonstrated that redistribution of high-density regions of histone 3 lysine 27 acetylation (H3K27ac), consistent with transcription-activating super-enhancer function, within condensates may contribute to platinum resistance in human ovarian cancer cell lines (*6*). These two observations suggest an important, dynamic link between transcriptional units and epigenetic marks in nuclear condensates with pharmacologic agents possessing condensate-specific proclivities. These studies further suggest that the modulation of condensate structure and function has critical roles in gene regulation and therapeutic response. This raises the question of whether the localization and efficacy of cisplatin and of other small-molecule therapeutics are driven by transcriptional and epigenetic regulation, by the structure of nuclear condensates, or by these two dynamic processes acting in concert.

Undoubtedly, there remain technical limitations to the ability to assess and elucidate the relationship between nuclear structure and specific perturbations, particularly in dynamic contexts, including malignant transformation and the response to specific therapies. As a cell responds to intracellular and extracellular signals, transcription factor activity is influenced by both chromatin conformation and other dynamic factors that have a critical role in these same biologic processes. Notably, a recent description of genetic alterations in transcription factors was shown to affect

Pharmacophore-specific condensates

Cisplatin and tamoxifen aggregate in transcriptional condensates, affecting their pharmacokinetics. Cisplatin freely diffuses through HP1-labeled heterochromatin and accumulates in MED1 condensates. The BRD4 inhibitor JQ.1 abrogates MED1 condensates and attenuates drug-induced platination. Tamoxifen displaces ERα from target genes, but mutations that induce tamoxifen resistance (mERα) shift affinity as tamoxifen is evicted from transcriptional condensates.



BRD4, bromodomain-containing protein 4; ERα, estrogen receptor α; HP1, heterochromatin protein 1; MED1, mediator of RNA polymerase II transcription subunit 1.

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. Email: leviner@mskcc.org

Published by AAAS

spatial localization to transcriptional condensates (*7*). Indeed, transcription-directed compartmentalization may have a critical role in regulating condensates, an observation best illustrated through the effects of herpes simplex virus type I in altering nuclear organization through highly accessible viral DNA binding sites and sequestration of RNA polymerase II (*8*).

Consistent with this hypothesis, intrinsically disordered regions within transcription factors can contribute to altered nuclear localization in human disease, as has been shown in syndactyly (fused fingers) driven by an intrinsically disordered region repeat expansion in homeobox protein 13 (HOXD13) (*7*). As was shown by Klein *et al.* with respect to pharmacophore-specific condensates, fluorescence recovery after photobleaching (FRAP)–based techniques were used to elucidate altered condensate formation in HOXD13 expansion syndactyly. Although FRAP cannot fully distinguish liquid diffusion from high-affinity protein structures (*9*), the relative contribution of liquid-liquid phase separation versus protein structured "hubs" remains an exciting and controversial question with respect to nuclear condensate formation and function.

These recent studies have dissected the inner workings of the nucleus, which are as beautiful and fascinating as much as they are still beyond complete understanding with respect to regulatory function and purpose. Enormous potential lies in the ability to manipulate these characteristics for therapeutic intervention. It is anticipated that further research will soon reveal the mechanistic importance of nuclear phase separation in cellular function, and the molecular consequences of linking spatial compartmentalization with nuclear function. ■

REFERENCES AND NOTES

1. C. P. Brangwynne, T. J. Mitchison, A. A. Hyman, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4334 (2011).
2. M. Feric *et al.*, *Cell* **165**, 1686 (2016).
3. I. A. Klein *et al.*, **368**, 1386 (2020).
4. A. Boija *et al.*, *Cell* **175**, 1842 (2018).
5. B. R. Sabari *et al.*, *Science* **361**, eaar3958 (2018).
6. Q. Ma *et al.*, *Cell Rep.* **31**, 107532 (2020).
7. S. Basu *et al.*, *Cell* **181**, 1062 (2020).
8. D. T. McSwiggen *et al.*, *eLife* 8, e47098 (2019).
9. D. T. McSwiggen *et al.*, *Genes Dev.* **33**, 1619 (2019).

ACKNOWLEDGMENTS

Studies supported by Memorial Sloan Kettering core facilities were funded in part by Memorial Sloan Kettering Cancer Center Support Grant/Core Grant P30 CA008748. A.D.V. is supported by National Cancer Institute career development grant K08 CA215317, the William Raveis Charitable Fund Fellowship of the Damon Runyon Cancer Research Foundation (DRG 117-15), and an EvansMDS Young Investigator grant from the Edward P. Evans Foundation. R.L.L. is on the supervisory board of Qiagen and is a scientific adviser to Loxo/Lilly, Zentalis, Imago, Mana Therapeutics, Auron, Ajax, Syndax, C4 Therapeutics, and Isoplexis. He receives research support from Prelude Therapeutics and has consulted for Celgene and Gilead.

10.1126/science.abc5318

ECOLOGY

Using information theory to decode network coevolution

Communication clashes shape the coevolution of insect-plant ecosystems

By **Ricard Solé**

Walking through a forest, you spot a colorful butterfly larva crawling and munching on a leaf—nothing unusual, just one scene in a calm ecological play. And yet, a massive "arms race" rages between plants and their herbivores (*1*, *2*), spurred by information and misinformation transfer. Chemical signals play the role of communicators in a channel that joins each pair of interacting partners. The mechanisms that drive coevolution of these chemically mediated webs have been an active area of research, but a satisfactory theory has yet to be established. On page 1377 of this issue, Zu *et al.* (*3*) describe a new application of information theory to coevolutionary dynamics in animal-plant networks.

Many scholars have explored the role of information in ecology and evolution (*4*) since the 1950s, in the wake of Claude Shannon's groundbreaking theory of information and communication (*5*). But the actual function of information, particularly in an evolutionary context, often has been obscured by insufficient data and the lack of a proper mapping of the links between information and fitness (*6*).

Zu *et al.* made use of plants' secondary metabolism (which forms metabolites not involved in plant growth or development) to couple two bipartite networks, namely the animal-plant (AP) one (i.e., who eats what) and the plant–volatile organic compounds (PV) one [i.e., what volatile organic compounds (VOCs) plants generate]. The authors gathered data from a tropical dry forest; insect larvae were collected from leaves, and trophic interactions were confirmed in the laboratory. The VOC repertoire was retrieved from leaves in the field with a method that characterizes the chemical profile around each leaf.

However, the way the authors built a communication channel between the AP and PV networks departed from Shannon's approach in a substantial way. The classical theory considers a sender and a receiver trying to communicate using a given code sent through a noisy channel (Morse code in the old telegraph system would be one example). For each signal sent, there is a probability that it will be misunderstood by the receiver. Not surprisingly, a great deal of standard information theory deals with finding ways of optimizing communication efficiency. However, in an ecological system in which interactions are highly asymmetric (one species eats the other), the needs at the ends of the (coevolving) communication channel are in clear conflict: Insects need to faithfully identify what leaves are edible, and plants need to avoid being identified as edible.

The aim of Zu *et al.* was to connect information-level descriptions of the two networks through a simple coevolutionary model that successfully solves the problem of defining fitness values for both plants and insects. The rules of the model simulate genetic mutations that influence pairwise interactions within the two bipartite networks. In the simulation, a given PV pair was chosen at random. With a certain probability, the connection between the plant and the chemical signal was added (if there was no link) or removed (if there was one), thus increasing or decreasing the VOC repertoire, respectively. A fitness value, F_P, was then computed using one of Shannon's entropic measures, which weights the uncertainty associated with chemical recognition by animals. If the new fitness value, F'_P, is larger than the original (before the PV choice), the change is accepted, and the VOC repertoire is expanded.

This feature of Zu *et al.*'s coevolutionary model is particularly notable, as communication networks usually are intended to increase channel efficiency through noise reduction (maximizing information transfer between sender and receiver). In their model, however, changes in a plant's VOC repertoire expanded the chemical diversity of volatiles that favor the confusion of the herbivores or, in Shannon's theory, that increase the herbivores' uncertainty about what plant choices they can make.

Institut de Biologia Evolutiva, Universitat Pompeu Fabra, Barcelona, Spain, and Santa Fe Institute, Santa Fe, NM, USA. Email: ricard.sole@upf.edu

Published by AAAS

Information between an herbivore (such as a caterpillar) and plant is exchanged through volatile chemical signals, a communication that drives coevolutionary arms races.

PHOTO: ANTONIO LÓPEZ-CARRETERO

The same mutation process was then applied to the herbivore-plant matrix, but now the chosen pairs were AP (insect-plant) duos. These pairs were also added or removed using a fitness measure. In this case, the selection pressure takes place in the opposite direction, toward reduction of the range of alternative chemical signals. Fitness is thus defined from another entropic measure that weights the improvement of insect specialization obtained from the animal-VOC matrix, which is the product of information transfer between the PV and the AP networks. In this way, two optimization constraints—plant VOC production and insect specialization—must be satisfied simultaneously.

With their new model, Zu *et al.* calculated several information measures from sets of interactions between lists of pairwise exchanges. The authors also extracted probabilities of interactions as well as so-called conditional probabilities (e.g., the probability that a given insect will recognize a specific plant). The sets of probabilities were used to estimate the efficiency of coding and decoding strategies. These measures and the predicted network structure matched very well with those estimated from the field data, including the dense set of connections in the PV data as well as the pattern of insect specialization in the AP set. These results were independent of initial conditions, which included both sparse and dense interaction webs. Such agreement supports the proposal that coevolutionary dynamics is at play and illuminates how VOCs influence the fitness of both plants and herbivores (*7*).

Both matter (biomass) and energy (metabolism) are key to our understanding of life. Information stored in DNA and signals transmitted between individuals are major driving forces in biology (*8*). In ecology, information theory has delivered testable predictions in the study of energy-flow networks (*9*) and in characterizing the ecological specialization of plant-pollinator networks (*10*). The theoretical framework presented by Zu *et al.* explicitly connects coevolutionary dynamics and network architecture. This framework can be applied to microbial VOC–based communication associated with plants (*11*) or the evolutionary dynamics of the rhizosphere, where a diverse range of signaling processes connect plants, fungi, and microorganisms (*12*).

In a general context, this formal approach to network coevolution could aid efforts to formulate a theory of ecosystem learning through coevolution (*13*). Furthermore, the opposing pressures that shape the evolution of language networks—which is closely related to Zu *et al.*'s modeling efforts—also include simultaneous minimization of communication costs and generate statistical distributions that reveal universal laws in language organization (*14*). ■

REFERENCES AND NOTES

1. L. M. Schoonhoven, J. J. A. van Loon, M. Dicke, *Insect-Plant Biology* (Oxford Univ. Press, 2005).
2. G. A. Howe, G. Jander, *Annu. Rev. Plant Biol.* **59**, 41 (2008).
3. P. Zu *et al.*, *Science* **368**, 1377 (2020).
4. M. I. O'Connor *et al.*, *Front. Ecol. Evol.* **7**, 219 (2019).
5. C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication* (Univ. of Illinois Press, 1949).
6. M. C. Donaldson-Matasci, C. T. Bergstrom, M. Lachmann, *Oikos* **119**, 219 (2010).
7. I. T. Baldwin, *Curr. Biol.* **20**, R392 (2010).
8. M. van Baalen, *Interface Focus* **3**, 20130030 (2013).
9. R. E. Ulanowicz, *Entropy* **21**, 949 (2019).
10. N. Blüthgen *et al.*, *Curr. Biol.* **17**, 341 (2007).
11. V. Bitas, H.-S. Kim, J. W. Bennett, S. Kang, *Mol. Plant Microbe Interact.* **26**, 835 (2013).
12. V. Venturi, C. Keel, *Trends Plant Sci.* **21**, 187 (2016).
13. D. A. Power *et al.*, *Biol. Direct* **10**, 69 (2015).
14. B. Corominas-Murtra, J. Fortuny, R. V. Solé, *Phys. Rev. E* **83**, 036115 (2011).

ACKNOWLEDGMENTS

R.S. is supported by the Generalitat de Catalunya and the Santa Fe Institute.

10.1126/science.abc6344

Published by AAAS

POLICY FORUM

EDUCATION

Understanding persistent gender gaps in STEM

Does achievement matter differently for men and women?

By **Joseph R. Cimpian[1], Taek H. Kim[1], Zachary T. McDermott[2]**

Gender gaps in science, technology, engineering, and math (STEM) college majors receive considerable attention, and it is increasingly recognized that not all STEM majors are equal in terms of gender disparities (*1*–*4*). For example, the male-to-female ratio among U.S. college majors in biology, chemistry, mathematics, and many other STEM fields is now about 1-to-1 (*2*, *5*), whereas in physics, engineering, and computer science (PECS), the ratio appears to have plateaued at about 4-to-1 (*2*, *4*, *5*). Here, we make two important contributions, showing (i) how gender relates to pursuit of a PECS degree throughout the achievement distribution and (ii) that student characteristics that predict PECS pursuit in the literature are not equally predictive of the gender gap throughout the achievement distribution. We find that a surprisingly large number of low-achieving men are majoring in PECS, relative to women, and this cannot be explained by an extensive set of student-level factors proposed in the prior literature (*6*–*10*). We can, however, explain the gender gap among high-achieving students. These patterns suggest that interventions to close the gender gap may work to attract high-achieving women; yet, something beyond these student factors may be attracting low-achieving men and repelling average- and low-achieving women, and without addressing those factors, it is unlikely that the PECS gender gap will fully close.

Despite historical emphasis on measuring whether gender gaps in STEM achievement exist and if there is greater male variability in achievement, there is surprisingly little research linking gender differences in achievement to pursuit of a college major [see supplementary materials (SM), section 1]. Work that links these factors accounts for average differences in STEM achievement and suggests that controlling for them does little to reduce the gender gaps in PECS, but it does not tell us where in the achievement distribution the PECS majors are coming from or if this differs for men and women.

We use new data from the U.S. Department of Education's High School Longitudinal Study of 2009 (HSLS:09), a nationally representative, longitudinal cohort of students who were in ninth grade during the 2009–2010 school year and were followed for 7 years. For several reasons, the HSLS:09 is the ideal dataset for examining who majors in PECS throughout the achievement distribution. First, it allows us to study a critical 7-year period, from the beginning of high school through the first few years of college when students declare majors (*3*, *11*). Second, its multiple measures of STEM achievement during high school allow for more comprehensive and robust measurement than most datasets. We combine these measures of STEM achievement into a composite STEM achievement variable with very high reliability (Cronbach's $\alpha = 0.86$; moreover, each individual measure yields a similar pattern of gender gaps; see fig. S1). Third, surveys conducted during the summer and fall of 2013 (after high school graduation for most) asked students about their college major intentions; then, they were asked 3 years later which major they were ultimately pursuing, allowing us to examine persistence in interest in a major and newfound interest in a major.

> **"...low-achieving men are both retained in and attracted to these fields during college..."**

The final analytic dataset contains 5960 students, with complete data on majors and STEM achievement, and multiply-imputed data for any missing covariates. All analyses use attrition-adjusted longitudinal sampling weights to ensure national representativeness and to adjust standard errors to account for the multistage survey sampling design. Detailed methods are in the SM. Although the data do not capture degree completion, they are likely a reasonable proxy, because research suggests that persistence to STEM degree is similar for both genders (see SM, section 2). The data reflect the United States, not international contexts.

GAPS DIFFER AT THE TOP VERSUS BOTTOM

Of all U.S. students attending a 4-year college, 23.8% of men and 5.5% of women pursued a PECS major, representing a substantial and significant gender gap on average ($p < 0.001$). When ranked from lowest to highest achieving, we see that males at and below the 1st percentile are majoring in PECS at the same rate as females at the 80th percentile of STEM achievement (see the figure, top). At every point in the STEM achievement distribution, men majored in PECS at higher rates than did women (all values of $p < 0.004$). The raw percentage-point gap in the top third of the achievement distribution is about twice the size of the gap in the bottom third (23 versus 12%, all values of $p < 0.001$; see Model 1 in the table). The size of these gaps is worth noting when thinking about the scale of the disparity throughout the distribution; however, percentage-point gaps do not factor in base rates, and thus, the gap represented in ratios is critical when examining representation. Importantly, the ratio at the bottom of the distribution is much larger than at the top. Overall, the male-to-female ratio in PECS is 4-to-1, whereas at the bottom decile, it is greater than 10-to-1 and at the top decile less than 2-to-1 (see the figure, top). This point is critical for considering retention in the PECS "pipeline" and should fundamentally shift how progress toward gender equity is evaluated.

Who majors in PECS can be disaggregated further by intention just after high school graduation to major in PECS: students who intended to major in PECS versus those who had different intentions (see the figure, middle and bottom). More males intended to major in PECS than did females (23.7 versus 5.2%, $p < 0.001$). Although persistence rates from post–high school intention to college major are similar overall (74 versus 75% for men and women, respectively; $p = 0.87$), these rates mask a substantial gender differential throughout the distribution, where females and males persisted at similar rates among the highest achievers but males persisted more among the lowest

[1]Department of Applied Statistics, Social Science, and Humanities, Steinhardt School of Culture, Education, and Human Development, New York University, New York, NY, USA. [2]Robert F. Wagner Graduate School of Public Service, New York University, New York, NY, USA.
Email: joseph.cimpian@nyu.edu

Published by AAAS

achievers (see the figure, middle). Indeed, among PECS-intenders who persisted, females averaged almost nine percentiles higher on the STEM composite measure (p = 0.004; table S1), because the male average was brought down by low achievers who persisted. This pattern of men's persistence being less sensitive to feedback on their actual performance is consistent with a general pattern found in other recent work (see SM, section 3). Importantly, males not previously intending to major in PECS join these majors at significantly higher rates than females at all points in the distribution (all values of $p < 0.008$; see the figure, bottom). Among non-PECS-intenders, males scoring at and below the 1st percentile were at least as likely to join PECS as females scoring above the 99th percentile who did not initially intend to major in PECS.

Supplemental analyses disaggregated PECS into physics only, engineering only, and computer science only (figs. S2 to S4), and common (although not identical) patterns were found, supporting the combining of majors for this main analysis.

EXPLANATION GAP AT THE BOTTOM

Are these gaps explained by other differences between men and women? Prior research suggests that women may choose PECS (and STEM, more generally) less often than men because they have lower confidence in their math abilities (*6*), have other options owing to a comparative advantage in English and/or reading (*7*), value societal goals and a work-life balance more than salary (*8*), take different high school STEM courses (*9*), and have different career aspirations (*10*).

Our ability to statistically account for the PECS gender gap with these and other variables depends on whether we look at low-, average-, or high-achieving students (see the table and fig. S5). Overall, our model that simultaneously adjusts for a range of student attributes explains almost all of the gender gap among high achievers, but not a substantial amount among low achievers. Among the top third of achievers, the average male-female gap goes from 23.0% (p < 0.001) in the unadjusted model to just 0.3% and no longer significant (p = 0.84) in the final adjusted model, for a total reduction of 98.6% [i.e., equal to (23.0% – 0.3%)/23.0%]. The rich set of covariates, however, does not do as well predicting the gaps among average- or low-achieving students. For example, the gap among the bottom third reduced only from 11.7 to 7.7% and remained significant (p < 0.001) after all adjustments (see the table). All patterns are robust to a wide range of sensitivity checks, including mean replacement for missing data, not using sampling weights, using cross-sectional weights, nonparametric models, and restricting the sample to students in colleges with engineering programs (figs. S6 and S7).

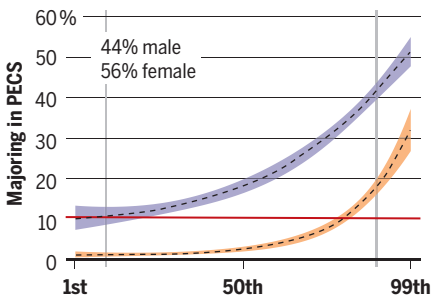
Who majors in PECS?

The shaded areas represent one standard error above and below the estimated means.

Males Females

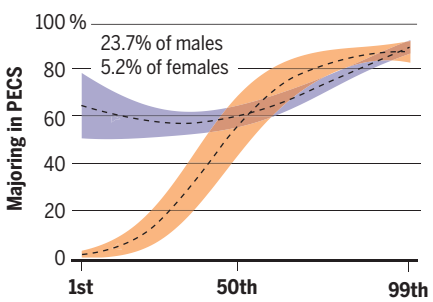
All students

Males scoring at or below the 1st percentile of all students' science, technology, engineering, and math (STEM) achievement major in physics, engineering, and computer science (PECS) at the same rate (10%, horizontal red line) as females scoring at the 80th percentile. The gap is significant at all levels of achievement. The male:female ratio in PECS is much larger in the 1st decile (10:1) than in the 10th decile (2:1) (vertical gray lines).



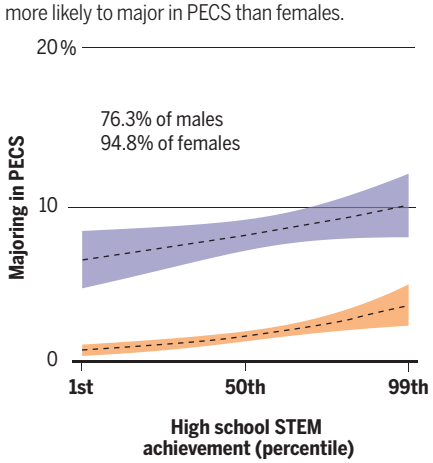
Intended PECS after high school

Among students who intended to major in PECS, higher-achieving males and females major in it at about the same rates, but lower-achieving males major in PECS at much higher rates than same-achieving females.



Did not intend PECS after high school

Among nonintenders, at each percentile, males are more likely to major in PECS than females.



NON-PECS AND RACE

We examine gender gaps in other STEM majors and race gaps in STEM as a point of comparison for our PECS-focused analyses (fig. S5 and tables S2 and S6 to S9. We find no significant gender gaps on average or throughout the achievement distribution in other STEM majors (table S11) after adjusting for all covariates. Similarly, there are no significant race gaps [Asian or white versus underrepresented racial minorities (URM)] in either PECS or other STEM, either on average or throughout the distribution after an extensive set of controls. It is also worth exploring gender-by-race patterns in PECS (*12*). The data revealed notable consistencies in gender gaps when looking just among Asian or white individuals or just among URM race groupings (figs. S8 and S9).

IMPLICATIONS

This new work reveals the critical importance of assessing the gender balance in PECS throughout the achievement distribution, rather than averages. Here, we discuss how this distributional approach (i) can transform how the field evaluates gender equity in the STEM pipeline and (ii) differentially affects how to target interventions to high versus average and low achievers. Ultimately, this approach suggests that student-level factors are complex and can vary across achievement levels. We urge the field to continue to look beyond student-level factors—toward discipline and societal ones—to further reduce the gender gap and raise the overall achievement of students in PECS.

IMPLICATIONS FOR THE STEM PIPELINE

Some have argued that the STEM pipeline from undergraduate to graduate school has stopped "leaking" because the proportion of females in STEM at the undergraduate level has remained constant to the graduate level (*1*, *13*). Our analyses suggest that looking at average persistence rates may obscure differential leaking throughout the achievement distribution. If PECS graduate programs selected from the highest-achieving students solely on the basis of merit, we should expect to see a considerably higher proportion of females in PECS graduate programs than in undergraduate programs. For example, if graduate programs select students on the basis of STEM achievement alone from the top decile, where the male-to-female ratio is less than 2-to-1, we should

GRAPHIC: X. LIU/*SCIENCE*

Published by AAAS

expect to see more than 33% females in graduate school, not 20%, which is the average during undergraduate school.

Thus, the apparent lack of a leaking pipeline on average found in previous studies (*1*, *13*) is not the appropriate standard for evaluating equity because it ignores that female PECS majors were initially higher achieving. Furthermore, evidence of differential leaking throughout the distribution could help reconcile seemingly disparate findings, such as a pipeline that no longer leaks on average (*1*, *13*) and male-dominated fields holding beliefs that deter females while sending welcoming signals to lower-achieving males for graduate education (*2*, *14*). It may be precisely through a male-favoring culture that graduate PECS programs keep the proportion of females at the same low rates as they are during undergraduate school rather than allowing them to rise, as would be expected given females' aforementioned higher competence.

Modeling the gender gap in PECS

		BY TERTILES OF HIGH SCHOOL STEM ACHIEVEMENT		
	OVERALL	LOW	AVERAGE	HIGH
MODEL 1: Unadjusted gap	18.2***	11.7***	17.2***	23.0***
MODEL 2: STEM composite adjustments	16.1***	10.3***	15.6***	22.3***
MODEL 3: Student attributes and attitudes	9.9***	9.5***	11.2***	9.0***
MODEL 4: Aspirations and intentions	5.5***	7.4***	5.8***	3.2*
MODEL 5: All adjustments	4.1***	7.7***	4.5**	0.3

All gaps are expressed in percentages. All estimates are average marginal effects, following logistic regressions predicting the major as a function of gender (Model 1), throughout the achievement distribution (Model 2, and subsequent models). Model 3 adjusts for factors of math and science interest, identity, self-efficacy, and utility, reported in grades 9 and 11, all interacted with achievement. Model 4 adjusts for college major intentions and occupational aspirations at age 30, reported in both grades 9 and 11. Model 5 adjusts for all variables listed above. See supplementary materials for details and for the underlying logistic regression output.
***$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.**

IMPLICATIONS FOR ACHIEVEMENT LEVELS

The differential ability of covariates to explain the PECS gender gap at different points in the achievement distribution has direct implications for how researchers and educators should intervene.

For high-achieving students, our analyses suggest that efforts to boost women's PECS aspirations and intentions (i.e., the strongest set of predictors)—such as role-model interventions and Girls Who Code (*4*, *15*)—may work well to eliminate the gender gap in pursuing PECS at the top of the distribution. That is, although each set of variables chipped away at the gender gap at the top of the distribution, none of them—alone or collectively—were as strong or consistent a predictor as prior intention to major in PECS and having aspirations for a PECS-related career (see the table, fig. S5A, and table S6). This clearly implies that to close the PECS-pursuit gender gap among the top achievers, the critical piece is in converting all other STEM-positive factors (e.g., confidence and interest) into actual aspirations and intentions.

Unfortunately, for average- and low-achieving students, our analyses also revealed that the field's understanding of how to get average- and low-achieving girls into the field is more limited. The same factors that explained the PECS gender gap at the top of the distribution collectively explained less and less of the gap further down the distribution. Thus, the interventions that may work for high-achieving girls are unlikely to work for lower-achieving ones. Moreover, focusing efforts on attracting high-achieving women to PECS may backfire: Whereas men have male peers to serve as role models throughout the achievement distribution, having only high-achieving role models for women may send signals to average- and low-achieving women that they do not have the necessary STEM skills to succeed (see SM, section 4).

Disaggregating PECS majors by achievement tiers raises a related question: Should PECS be attracting and retaining low-STEM-achieving students (of any gender), or is there an alternative pathway to meet the PECS labor market demands with more competent students? Students in the lowest third of achievement have mean math SAT scores and high-school STEM grade point averages that are well below average, especially for these math-intensive fields. About 90% of the students majoring in PECS from this tier are men, and we have shown that low-achieving men are both retained in and attracted to these fields during college (see the figure, middle and bottom, and figs. S3 and S4). Thus, the question about whether PECS wants low-achieving students is inextricably intertwined with a different question: Why are men—and not women—with low achievement more drawn to these fields? Although we have shown that this question cannot be answered by gender differences in a wide range of student-level factors—including self-efficacy, math or science identity, and earlier aspirations—it is possible that the masculine culture of these fields and the gender stereotypes attached to PECS (*2*, *4*, *11*, *14*, *15*) may lead to retaining less-qualified men over more-qualified women. Thus, it is important for PECS fields to break down the barriers for women to enter the field and, in doing so, raise the overall quality of the students entering these fields.

All of this suggests that, although role-model interventions, coding camps, and fostering strong peer networks are worth pursuing, there are real obstacles to achieving gender parity in PECS that transcend these interventions, which may work only for a select group of high-achieving women. To develop effective interventions for gender equity and raise the overall quality of students in PECS, it is critical to recognize that the gender imbalance varies throughout the achievement distribution and may be perpetuated by male-favoring cultures that disproportionately attract low-achieving men to the field. ■

REFERENCES AND NOTES

1. S. J. Ceci, D. K. Ginther, S. Kahn, W. M. Williams, *Psychol. Sci. Public Interest* **15**, 75 (2014).
2. S. Cheryan, S. A. Ziegler, A. K. Montoya, L. Jiang, *Psychol. Bull.* **143**, 1 (2017).
3. C. Riegle-Crumb, B. King, E. Grodsky, C. Muller, *Am. Educ. Res. J.* **49**, 1048 (2012).
4. C. Hill, C. Corbett, A. St. Rose, "Why so few? Women in science, technology, engineering, and mathematics" (Report, American Association of University Women, 2010).
5. National Science Board, "2018: Science & engineering indicators" (NSB-2018-1, National Science Foundation, 2018).
6. L. J. Sax, M. A. Kanny, T. A. Riggers-Piehl, H. Whang, L. N. Paulson, *Res. High. Educ.* **56**, 813 (2015).
7. T. Breda, C. Napp, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15435 (2019).
8. A. B. Diekman, M. Steinberg, E. R. Brown, A. L. Belanger, E. K. Clark, *Pers. Soc. Psychol. Rev.* **21**, 142 (2017).
9. N. Friedman-Sokuler, M. Justman, *Econ. Educ. Rev.* **53**, 230 (2016).
10. S. L. Morgan, D. Gelbgiser, K. A. Weeden, *Soc. Sci. Res.* **42**, 989 (2013).
11. C. M. Ganley, C. E. George, J. R. Cimpian, M. B. Makowski, *Am. Educ. Res. J.* **55**, 453 (2018).
12. The National Academies of Sciences, Engineering, and Medicine, "Promising practices for addressing the underrepresentation of women in science, engineering, and medicine: Opening doors (Consensus study report, The National Academies Press, 2020).
13. D. I. Miller, J. Wai, *Front. Psychol.* **6**, 1 (2015).
14. S.-J. Leslie, A. Cimpian, M. Meyer, E. Freeland, *Science* **347**, 262 (2015).
15. N. Dasgupta, J. G. Stout, *Policy Insights Behav. Brain Sci.* **1**, 21 (2014).

ACKNOWLEDGMENTS

We thank A. Cimpian, Y. Copur-Gencturk, S. Lubienski, M. Makowski, M. Perry, L. Stiefel, and J. Timmer for helpful comments. Z.T.M. is funded through an Institute of Education Sciences (IES) Predoctoral Interdisciplinary Research Training grant to New York University (R305B140037). Although HSLS:09 has a public-use version of the dataset, some variables used for the analysis are only available in a restricted-use dataset from the U.S. Department of Education. For information on restr Sicted-use licenses, see https://nces.ed.gov/pubsearch/licenses.asp. All Stata code and an ancilliary dataset can be found at https://osf.io/6pma7.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1317/suppl/DC1

10.1126/science.aba7377

Published by AAAS

BOOKS *et al.*

Biologist Nancy Hopkins used a data-driven approach to document subtle sexism at MIT in the early 1990s.

SCIENCE LIVES

Creating a culture of change

Urging action, a new film paints a harrowing portrait of female scientists' experiences in academia

By **Donna Riley**

Sweeping in scope yet intimately compelling, *Picture a Scientist* tells the stories of three female scholars, revealing the systemic and structural nature of gender discrimination and harassment in academic science. The film shows how intersections of sexism and racism shape experiences differently for white women and for women of color and how implicit bias both generates inequity and prevents us from noticing it. It reveals injustices ranging from subtle slights and salary gaps to bullying and physical assault. And it reminds us of the power of women's collective action; the value of social science research in analyzing and responding to inequality; and the importance of allies and advocates, especially in university leadership.

Picture a Scientist
Ian Cheney and Sharon Shattuck, directors
Uprising LLC, 95 minutes.
Streaming via select theaters 12–26 June 2020.

Biologist Nancy Hopkins shows what convivial conspiracies can hatch over lunch with a colleague, when one dares to act. In the early 1990s, convinced that male colleagues at the Massachusetts Institute of Technology (MIT) were afforded more lab space than their female counterparts, she took a tape measure to her own lab and to the lab spaces of male colleagues in the dead of night, carefully documenting the disparity. With this data-driven approach, Hopkins gained the support of her female colleagues and upper-level administrators and went on to lead a more comprehensive study investigating salary gaps, access to day care, and other inequities in need of redress (*1*).

But the sexism Hopkins experienced during her career was not always subtle. In the film, she recounts a shocking groping incident that occurred when she was an undergraduate presenting her work to a visiting colleague. Stunned by the encounter and certain that no one would believe her over her perpetrator, she resolved to act as though nothing unusual had occurred, answering the Nobelist's questions without mention of the offense.

Sexual harassment is more often a "put-down" than a "come-on," explains anthropologist Kate Clancy in the film's narration, as geoscientist Jane Willenbring reveals how she was bullied, humiliated, and assaulted while conducting fieldwork in Antarctica as a graduate student in 1999. After nearly two decades of silence, she felt compelled to speak out against the perpetrator when her young daughter began expressing scientific aspirations of her own. Willenbring's Title IX complaint resulted in her perpetrator's dismissal, due in part to the supporting testimony of a bystander, graduate student Adam Lewis. One wonders why Lewis did not intervene at the time, but as he shares his epiphanies and takes gentle correction from his colleague, we see how true allies are formed.

What a ponderous thing it is to hold an esteemed scientist accountable for unprofessional behavior; so much has to align to achieve any semblance of justice. For every story, there are thousands left untold; academia is haunted by those who cannot come forward, those who are not believed, and those who are dismissed as ill-suited for the discipline.

In chemist Raychelle Burks's story, we encounter the toll of day-to-day slights, underestimations, and the simultaneous invisibility and hypervisibility that women of color experience in the academy. As Burks recounts how a colleague challenged her right to park in the faculty lot, we see how such experiences can accrue over one's career, taking time and energy away from science. Psychologists Mahzarin Banaji and Corinne Moss-Racusin reveal how our implicit biases prevent us from associating women with excellence in science, generating inequity that slips by unnoticed unless we are vigilant.

Picture a Scientist is a world-class documentary from an experienced, award-winning crew that tells harrowing truths without sugarcoating, sensationalizing, or objectifying the film's subjects. If I must have a quibble, it is that the film's treatment of race is limited to black and white and its treatment of gender is too binary. Other axes of difference, such as disability, class, or sexual orientation, go largely unaddressed. Still, the film leaves openings to discuss these omissions and, more importantly, compels us to action. We might ask ourselves: What data could we gather on our own campuses? How do we become the accomplices of change-seeking colleagues? Can we muster the courage to share our own stories or hold someone accountable rather than looking the other way?

Picture a Scientist will be available to stream via select U.S. theaters from 12 to 26 June 2020. Invite campus leaders to attend the virtual premiere, and then host a screening and organizing session (*2*). This film could be the vehicle that moves us, as Clancy advocates, "away from a culture of compliance and towards a culture of change." ■

The reviewer is the Kamyar Haghighi Head of the School of Engineering Education, Purdue University, West Lafayette, IN 47907, USA. Email: riley@purdue.edu

REFERENCES AND NOTES

1. A Study on the Status of Women Faculty in Science at MIT (1999); http://web.mit.edu/fnl/women/women.html.
2. www.pictureascientist.com/#host-section

10.1126/science.abd0344

PHOTO: UPRISING LLC

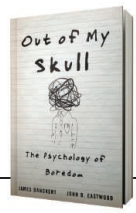
Published by AAAS

PSYCHOLOGY

What to do with our days

When wielded correctly, boredom can be a powerful tool

Out of My Skull: The Psychology of Boredom
James Danckert and John D. Eastwood
Harvard University Press, 2020. 288 pp.



By **Erin Westgate**

On 13 May 1982, Valentin Lebedev was launched into space, where he would spend a record-breaking 211 days. It took only a week for the astronaut to confront a surprising problem: He was bored. Lamenting the "drab routine" and tedious "busy work," his space diary documents his struggles and his intense anticipation of a video camera that he hoped would relieve his boredom.

Although we may not be adrift in space, boredom is often inescapable these days, as social distancing guidelines keep many of us at home. The key to mastering boredom, write John Eastwood and James Danckert in their intriguing new book *Out of My Skull*, is understanding its purpose. They argue that boredom is an important signal that "we need to act to fully realize our potential."

The same conditions can be experienced as either boring or deeply engaging, depending on our internal state.

PHOTO: ISTOCK.COM/PATRICKHEAGNEY

Eastwood, a clinical psychologist at York University, and Danckert, a cognitive neuroscientist at the University of Waterloo, introduce boredom as "the uncomfortable feeling of wanting, but being unable, to engage in satisfying activity." Defining it as a "feeling of thinking," they argue that boredom stems from four situational sources: monotony, constraint, lack of purpose, and lack of optimal challenge. When these contexts interact with individual internal states, such as motivation and attentional capacity, the same conditions can be experienced as either boring or deeply engaging.

Drawing a rigorous distinction between the feeling of being bored ("boredom") and the tendency for some people to be bored more easily than others ("boredom proneness"), Eastwood and Danckert systematically review empirical evidence in an attempt to determine whether boredom is good or bad, ultimately concluding that it is neither. Just like play, they argue, boredom serves important functions across species, guiding behavior away from inaction, toward novel and potentially rewarding outcomes.

Bored minks, for instance, behave very much like bored humans in lab experiments. When housed in unenriched cages, they become curious about exploring any and all new objects, from plastic water bottles to a moving toothbrush (which, the authors assure us, a mink finds quite exciting). And, like humans who are willing to self-administer painful electric shocks when bored (*1*), minks, too, show as much interest in aversive items, such as the smell of a predator, as they do in more positive ones.

Evidence that boredom propels exploration and learning also emerges from computer science. Programming artificial intelligence with the capacity for boredom provides a self-regulatory feedback loop that encourages optimal challenge and discourages programs from persisting at tasks that are either too complex to be feasible or too easy to offer new opportunities for insight. Indeed, when pitted against each other in computational models, it is boredom, not curiosity, that drives learning agents to make the most discoveries.

What does this mean in an age when we can easily suppress boredom with a few swipes on a smartphone? Eastwood and Danckert urge readers to heed the possibility that technology may offer a tempting but temporary solution. Brushing boredom away with "solutions," such as social media, does nothing to solve the lack of purposeful engagement that gives rise to boredom in the first place. While skeptical of claims that boredom is particularly harmful or beneficial—duly noting that experimental evidence is still in its infancy—they argue that we should take it seriously, by making room for it in our lives and responding appropriately when it arises.

One clue to remedying boredom lies in its emotional opposites: interest, enjoyment, and optimal "flow" experiences. By occupying our minds with activities we care about and finding ways to effectively pursue our desires and goals, we can act not only to address boredom as it occurs but also to prevent its occurrence in the first place.

Although Eastwood and Danckert take pains to clarify that they do not consider boredom to be an emotion, their approach is highly consistent with modern theories in social psychology and affective science that define emotions as "situated affective reactions," whose purpose is to provide information about whatever is in mind at the moment, our own thinking included (*2*–*5*). Drawing from this broader literature on emotion, and related work on the psychology of meaning and motivation, would have enriched the book's already broad approach to understanding what boredom is, why we experience it, and what happens when we do. Overall, however, Eastwood and Danckert's synthesis is an engaging and timely read that is anything but boring. ■

The reviewer is at the Department of Psychology, University of Florida, Gainesville, FL 32611, USA.
Email: erinwestgate@ufl.edu

REFERENCES AND NOTES

1. T. D. Wilson *et al.*, *Science* **345**, 75 (2014).
2. S. Schachter, J. Singer, *Psychol. Rev.* **69**, 379 (1962).
3. L. F. Barrett, *How Emotions Are Made: The Secret Life of the Brain* (Houghton Mifflin Harcourt, 2017).
4. A. Ortony, G. L. Clore, A. Collins, *The Cognitive Structure of Emotions* (Cambridge Univ. Press, 1988).
5. G. L. Clore, J. R. Huntsinger, *Trends Cogn. Sci.* **11**, 393 (2007).

10.1126/science.abc1024

Published by AAAS

LETTERS

Edited by **Jennifer Sills**

Academic societies' role in curbing police brutality

The shocking death of George Floyd is unfortunately not a singular event in the United States. Eric Garner, Michael Brown, Freddie Gray, Amadou Diallo (*1*), and Breonna Taylor (*2*) are just a few examples from a long list of unarmed Black people who have died at the hands of the police. This pattern of violence is deeply rooted in the history of Black-White relations in the United States and the failure of the leaders of this country to deal with systemic racism. As professional physicists, we suggest that professional science and humanities organizations such as the American Physical Society (APS) take action by refusing to hold conferences in cities where police brutality takes place.

The APS and many other professional societies host large meetings in cities across the United States. APS cancelled a 2018 meeting in Charlotte, North Carolina, to oppose a law that discriminated against the LGBT community (*3*). The American Association for the Advancement of Science (the publisher of *Science*) has also selected cities for meetings based on their record of civil rights, including excluding the "Jim Crow" South from consideration and supporting states that ratified the Equal Rights Amendment (*4*). In this spirit, professional societies should protect Black people by meeting only in cities that implement evidence-based policies to limit the use of force by police. Academic societies should also assess cities' responses to police brutality, such as the transparency of communication and whether offending officers have been fired and charged with a crime. Social scientists have researched policies that can reduce police violence [e.g., (*5*)], and we encourage science societies to use this evidence to develop a set of criteria.

Economic pressure has historically been an effective strategy in bringing about social change. A decision to stop doing business in cities with inadequate protective policies would be similar to the divestment movement in South Africa in the 1980s (*6*), which was pivotal in ending apartheid. It would also be in keeping with Dr. Martin Luther King Jr.'s boycott of businesses in the South that refused to hire Black people (*7*). Academic societies should encourage positive change by sending a powerful message: Cities that reduce police brutality through egalitarian reforms will yield substantial financial benefits.

By using their power of the purse to oppose racist policing, professional societies would also be protecting their minority members, who might, for example, decide during a meeting to take a run through city streets while being Black. The ongoing protests demonstrate that established norms of social and race relations are changing. It is time for professional societies to live up to their charge of positively affecting society by taking a stand against police brutality in deed as well as in word.

Philip Phillips* and Michael B. Weissman
Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.
*Corresponding author. Email: dimer@illinois.edu

REFERENCES AND NOTES

1. D. Funke, T. Susman, "From Ferguson to Baton Rouge: Deaths of black men and women at the hands of police," *Los Angeles Times* (2016).
2. E. Haines, "Family seeks answers in fatal police shooting of Louisville woman in her apartment," *The Washington Post* (2020).
3. APS, "Discriminatory North Carolina restroom law prompts major scientific society to move annual meeting" (2016); www.aps.org/newsroom/pressreleases/08042016.cfm.
4. S. Malcolm, *Science* **349**, 671 (2015).
5. M. Pryor, K. S. Buchanan, P. A. Goff, *Annu. Rev. Law Soc. Sci*. 10.1146/annurev-lawsocsci-101518-042633 (2020).
6. G. Gethard, "Protest divestment and the end of apartheid," *Investopedia* (2019).
7. S. Cohen, "Why the Woolworth's sit-in worked," *Time* (2015).

10.1126/science.abd1932

Pandemics' historical role in creating inequality

In her News Feature "An unequal blow" (15 May, p. 700), L. Wade explores how past pandemics disproportionately affected marginalized groups. It is important to recognize that in addition to exacerbating marginalization, pandemics set the stage for inequalities that have persisted for centuries (*1*). Historically, pathogens introduced by European conquistadors and colonists contributed to Indigenous population collapses throughout the Americas, facilitating colonial exploitation (*2*). Abrupt declines in the health and size of Indigenous populations, socio-cultural dislocation, and opportunistic grabs for resources became a strategy used by successive colonial governments (*3*). As scientists, we have a role in dismantling this colonial playbook to prevent coronavirus disease 2019 (COVID-19) from further marginalizing Indigenous communities.

Today's Indigenous peoples are facing a pandemic crisis. In North America, the Navajo Nation battles the virus across a vast region with irregular access to health services as a result of insufficient federal funding (*4*, *5*). The United States relies on agricultural laborers from Central America to maintain its supply chains but is deporting contagious workers (*6*, *7*), thereby seeding outbreaks in Mexico, where the disenfranchisement of 22 million Indigenous people makes the spread of disease difficult to control (*8*). In the Brazilian Amazon, Manaus is a national COVID-19 hotspot. The health system is overwhelmed, and Indigenous people previously displaced by the local government struggle to avert a 21st-century demographic collapse while fighting seizure of their lands (*9*).

Solving the crisis demands more than overhauling health infrastructure; it requires changing power relations to redress structural inequalities borne of colonialism. The path forward involves listening to Indigenous communities, promoting human-centered investment in sustainable development, and enforcing treaties and Indigenous land rights. For scientists, COVID-19 does not justify perpetuating extractive practices or pressuring communities to engage in research (*10*). Acknowledging Indigenous sovereignty is now more vital than ever. We must seize the opportunity this pandemic presents for societal transformation: Only through Indigenous self-determination, collaboration, and research partnerships rooted in

PHOTO: MICHAEL DANTAS/AFP/GETTY IMAGES

Published by AAAS

COVID-19 victims are buried daily in this cemetery in Manaus, Brazil, where the health system has been overwhelmed.

mutual respect can the Americas write a new playbook that fulfills the often stated but seldom realized ideals of equality for all.

Liliana M. Dávalos[1]*, Rita M. Austin[2,3,4], Mairin A. Balisi[5,6], Rene L. Begay[7], Courtney A. Hofman[2,3,8], Melissa E. Kemp[9], Justin R. Lund[2,3], Cara Monroe[2,3], Alexis M. Mychajliw[2,3,5], Elizabeth A. Nelson[10,11,12], Maria A. Nieves-Colón[13], Sergio A. Redondo[14], Susanna Sabin[15], Krystal S. Tsosie[16,17], Joseph M. Yracheta[17,18]

[1]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA. [2]Department of Anthropology, University of Oklahoma, Norman, OK 73019, USA. [3]Laboratories of Molecular Anthropology and Microbiome Research, University of Oklahoma, Norman, OK 73019, USA. [4]Department of Anthropology, National Museum of Natural History, Washington, DC 20560, USA. [5]Department of Rancho La Brea, Natural History Museum of Los Angeles County, Los Angeles, CA 90007, USA. [6]Department of Life and Environmental Sciences, University of California Merced, Merced, CA 95343, USA. [7]Centers for American Indian and Alaska Native Health, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA. [8]Department of Anthropology, National Museum of Natural History, Washington, DC 20560, USA. [9]Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA. [10]Department of Sociology and Anthropology, University of Texas at Arlington, Arlington, TX 76019, USA. [11]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [12]Institut für Naturwissenschaftliche Archäologie, Eberhard Karls Universität Tübingen, 72074 Tübingen, Germany. [13]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85281, USA. [14]Department of Biology, Stanford University, Stanford, CA 94305, USA. [15]Center for Evolution and Medicine, Arizona State University, Tempe, AZ 85281, USA. [16]Interdisciplinary Graduate Program, Vanderbilt University, Nashville, TN 37235, USA. [17]Native BioData Consortium, Eagle Butte, SD 57625, USA. [18]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA.
*Corresponding author.
Email: liliana.davalos@stonybrook.edu

REFERENCES AND NOTES

1. T. Hansen, J. Yracheta, "Indigenous migrant children at higher risk for flu-related deaths, yet denied basic flu shot," *Cultural Survival* (2019).
2. A. W. Crosby, *The Columbian Exchange: Biological and Cultural Consequences of 1492* (Greenwood Publishing Group, Westport, CT, 1973).
3. J. R. McNeill, *Mosquito Empires: Ecology and War in the Greater Caribbean, 1620–1914* (Cambridge University Press, Cambridge, UK, 2010).
4. Navajo Nation Department of Health, "Dikos Ntsaaígíí-19 (COVID-19)" (2020); www.ndoh.navajo-nsn.gov/COVID-19.
5. U.S. Commission on Civil Rights, "Broken Promises: Continuing Federal Funding Shortfall for Native Americans" (2018); www.usccr.gov/pubs/2018/12-20-Broken-Promises.pdf.
6. M. Dorning, J. Skerritt, "Every Single Worker Has Covid at One U.S. Farm on Eve of Harvest," *Bloomberg* (2020).
7. K. Sieff, N. Miroff, "U.S. is deporting infected migrants back to vulnerable countries," *The Washington Post* (2020).
8. L. Castellanos, "México atropella los derechos de Indígenas y defensores ante el coronavirus" *The Washington Post* (2020) [in Spanish].
9. I. Amigo, "Evicted Indigenous people in Manaus struggle to stay safe amid COVID-19 crisis," *Mongabay* (2020).
10. K. G. Claw *et al.*, *Nat. Commun.* **9**, 2957 (2018).

10.1126/science.abc8953

Recent immigrants at increased pandemic risk

In her News Feature "An unequal blow" (15 May, p. 700), L. Wade discusses the vulnerability of poor people and members of minorities during both past and present pandemics. However, the article overlooks recent immigrants, who have also been particularly vulnerable to pandemics in the past and are threatened now by the coronavirus disease 2019 (COVID-19) pandemic.

Human migration was also taking place during epidemics of the past. During the Black Death, some of those who died in London originated from distant portions of Britain (*1*), and rural refugees in Cairo, Egypt, exhibited a higher susceptibility to death (*2*). During plagues of the late preindustrial era in Milan, Italy, the death rate was higher among poor people originating from surrounding villages (*3*). In Dijon, France, during early recurrences of the Black Death, poor newcomers were more likely to die than were poor people who were long established in the city or who had recently settled but were members of established families (*4*).

The COVID-19 pandemic poses similar risks. In the U.S. states with newer Latino communities, Latinos have tested positive at higher rates than people with comparable income. In contrast, in some states with long-established Latino communities, Latinos are getting sick at rates similar to other population groups (*5*).

Although the comparative importance of frailty factors accounting for the present situation is a matter of debate, socio-economic and environmental factors are essential (*6*). This is consistent with the long-term historical perspective: During preindustrial European plagues, the vulnerable short-range migrants likely differed from the host population only in housing and working conditions. As past and present data demonstrate, recent immigrants should be included in discussions of risks to marginalized groups.

Pierre Galanaud[1]* and Anne Galanaud[2]

[1]UMR-996, Inflammation, Microbiome, and Immunosurveillance, Inserm, Université Paris-Saclay, Clamart, France. [2]Université de Bourgogne-Franche-Comté, Besançon, France.
*Corresponding author.
Email: pierre.galanaud@universite-paris-saclay.fr

REFERENCES AND NOTES

1. E. J. Kendall *et al.*, *Am. J. Phys. Anthropol.* **150**, 210 (2013).
2. S. Borsch, T. Sabraa, *Ann. Démograph. Hist.* **2**, 63 (2017).
3. S. K. Cohn, G. Alfani, *J. Interdisc. Hist.* **31**, 177 (2007).
4. P. Galanaud, A. Galanaud, P. Giraudoux, H. Labesse, *PLOS ONE* **15**, e0226420 (2020).
5. M. Jordan, R. A. Oppel Jr., "For Latinos and Covid-19, doctors are seeing an 'alarming' disparity," *The New York Times* (2020).
6. N. Bhala *et al.*, *Lancet* **395**, 1673 (2020).

10.1126/science.abd1098

Published by AAAS

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

GEOPHYSICS

Seismic swarms show the structure

Faults responsible for earthquakes are idealized into two dimensions, despite fault zones being complicated, three-dimensional structures. Ross *et al.* used machine learning to find 22,000 seismic events near Cahuilla, California, during a seismic swarm. They used the locations and sizes of these events to show how the complex structure of the fault interacted with natural fluid injections from below. The authors' methods highlight the complexities of one fault and suggest a way to characterize other faults around the world. —BG *Science*, this issue p. 1357

A line marks the site of an ancient lake in a fault-rich region near Salton City, California, not far from the site of a recent seismic swarm induced by fluid injection from a deep reservoir.

CORONAVIRUS

Promising antiviral protease inhibitors

With no vaccine or proven effective drug against the virus that causes coronavirus disease 2019 (COVID-19), scientists are racing to find clinical antiviral treatments. A promising drug target is the viral main protease M^{pro}, which plays a key role in viral replication and transcription. Dai *et al*. designed two inhibitors, 11a and 11b, based on analyzing the structure of the M^{pro} active site. Both strongly inhibited the activity of M^{pro} and showed good antiviral activity in cell culture. Compound 11a had better pharmacokinetic properties and low toxicity when tested in mice and dogs, suggesting that this compound is a promising drug candidate. —VV

Science, this issue p. 1331

METALLOENZYMES

Delicate dance becomes a ballet

The enzyme nitrogenase uses adenosine triphosphate and several unusual iron-sulfur cofactors to pump electrons into typically inert dinitrogen (N_2), providing protons along the way. Previous work has shown that sulfur atoms in the iron-molybdenum cofactor (FeMoCo) are labile and suggests that replacement of one of the sulfurs by N_2 is integral to the mechanism of N_2 binding and reduction. Through the elimination of excess reducing agent during preparation, Kang *et al*. determined structures of Mo-nitrogenase in a resting conformation. Unexpectedly, they found that all three sulfurs at the outer edge of FeMoCo appear to be labile, with one subunit even having two of three sulfurs replaced by light, diatomic ligands. Biochemical and spectroscopic data indicate that the protein is active, holds tightly bound N_2, and is in the expected oxidation state. These results may prompt a reassessment of the possible mechanisms of N_2 reduction and the role of dynamic belt ligands in FeMoCo. —MAF

Science, this issue p. 1381

METALLURGY

Strong and tough steel

Ultrahard materials often do not have similarly impressive fracture toughness. Liu *et al*. discovered a processing route for medium manganese steel alloy in which ultrahigh strength accompanies high fracture toughness. The steel relies on both transformation-induced plasticity and delamination toughening to boost the fracture properties. The steel is composed of less expensive elements, making it

PHOTO: DAVID MCNEW/GETTY IMAGES

Published by AAAS

a potentially inexpensive material attractive for structural applications. —BG

Science, this issue p. 1347

CANCER THERAPY

Speedy screen for tumor therapies

Cell death screens using patient biopsies could be used to identify effective treatments for cancerous solid tumors. However, it takes several days to obtain results, meaning that the cells guiding treatment decisions may become molecularly distinct from those in patients. Bhola *et al*. developed a high-throughput method that, within 24 hours, identifies drugs that initiate cell death programs in tumor cells from freshly isolated patient biopsies. This method identified drug therapies that shrank breast and colon tumors in mice, but which would not have seemed as promising when screened in cells cultured for several days. —LKF

Sci. Signal. **13**, eaay1451 (2020).

INFLUENZA

Resistance to influenza antibodies

Broadly neutralizing human antibodies (bnAbs) to the stem of hemagglutinin (HA), a trimeric glycoprotein found on the surface of influenza viruses, are valuable therapeutics and can guide the development of universal influenza vaccines. For their use in therapy development, it is important to understand the extent to which HA stem variants with resistance to bnAbs can develop. Wu *et al*. used saturation mutagenesis combined with next-generation sequencing to systematically search for resistance mutations to prototypic bnAbs in two influenza subtypes, H3 and H1. They found that the genetic barrier to resistance to stem bnAbs was low for the H3 subtype but higher for the H1 subtype. The ability of H3 to develop resistance to bnAbs presents a challenge in the development of a universal influenza vaccine. —VV

Science, this issue p. 1335

BIODIVERSITY CHANGE

Land-use change and forest biodiversity

Land-use change by humans, particularly forest loss, is influencing Earth's biodiversity through time. To assess the influence of forest loss on population and biodiversity change, Daskalova *et al*. integrated data from more than 6000 time series of species' abundance, richness, and composition in ecological assemblages around the world. Forest loss leads to both positive and negative responses of populations and biodiversity, and the temporal lags in population and biodiversity change after forest loss can extend up to half a century. Land-use change precipitates divergent population and biodiversity change. This analysis has consequences for projections of human impact, ongoing conservation, and assessments of biodiversity change. —AMS

Science, this issue p. 1341

ELECTROCHEMISTRY

Cutting it close for radical coupling

In principle, electrochemistry is an ideal method for radical coupling: One precursor oxidized at the anode pairs up with a counterpart that has been reduced at the cathode. The trouble is that either or both coupling partners might not stay stable long enough to meet in the middle. Mo *et al*. resolved this issue by closely spacing the electrodes in a microfluidics platform (see the Perspective by Liu *et al*.). They showcase coupling of dicyanobenzene as the cathodic radical precursor with a variety of oxidatively generated partners. —JSY

Science, this issue p. 1352; see also p. 1312

IN OTHER JOURNALS

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

Scanning transmission electron micrograph of wire-shaped morphology in BiSeI.

IMAGE: WANG *ET AL.*

THERMAL CONDUCTIVITY

Simple structures to slow heating

Materials with low thermal conductivity are important as barrier coatings and for other applications. Wang *et al.* found a bismuth selenohalide, BiSeI, with an extremely low thermal conductivity over a range of temperatures. The authors attributed their observations to the simple crystal structure of the material, along with other features that inhibited thermal transport. The characteristics that give bismuth selenohalide a low thermal conductivity could constitute a template for finding other materials that are good thermal barrier coatings or those that could be used for thermoelectrics. —BG

Sci. China Mater. 10.1007/s40843-020-1407-x (2020).

CORAL REEFS

Rot in the national bedrock

Coral reefs face numerous threats in today's changing environment. For low-lying atoll countries built on reef systems, the consequences can be catastrophic. The best-recognized threat to coral is climate change, but not all damage is outside local control. Local pollution seems to be taking a considerable toll in Tuvalu, a country in the South Pacific composed of several coral reef atolls. Nakamura *et al.* examined the composition of Tuvalu coral cores spanning decades. Since the 1990s, annual growth layers have been punctuated by black bands containing heavy metals, notably redox iron, and the genetic remnants of anaerobic bacteria. The authors

Published by AAAS

FORESTRY

Genetic variation in boreal conifers

Boreal forests are suffering from the effects of climate change. Trees take decades to grow, and it is difficult to test their adaptive capacity. Depardieu *et al.* took advantage of a common garden experiment set up in 1979 in which tree seeds from 43 geographic origins were planted in a single location in Quebec, Canada. Tree-ring data from white spruce cores in 2006 showed growth responses to droughts in 1997, 2001–2002, and 2005. Trees from locations with drier climates displayed alterations in xylem anatomy and better recovery after drought than trees originating from more humid locations. These variations may reflect genetic adaption to local climate conditions in this widely distributed northern conifer. This information will be valuable in efforts to offset drought sensitivity in reafforestation planning. —PJH

New Phytol. **227**, 427 (2020).

White spruce trees (*Picea glauca*) native to the boreal forests of North America show genetic adaptation to the local environment.

infer that the black bands are produced by annually occurring anoxic conditions caused by pollution-driven algal overgrowth and coral suffocation. Such direct evidence of local conditions eroding the very bedrock of a country may provide enough motivation for change. —SNV

Sci. Rep. **10**, 7338 (2020).

CELL BIOLOGY

Monitoring mitophagy

Mitophagy is the autophagic process by which mitochondria are degraded in lysosomes. It removes dysfunctional mitochondria that accumulate in a variety of human diseases. New tools are needed to quantify autophagy and mitophagy in living cells. Katayama *et al.* have developed a signal-retaining fluorescent autophagy indicator (SRAI). The signal is not affected by delivery to the low-pH lysosomal environment and can be used in fixed as well as live samples. In a mouse model of Parkinson's disease, the indicator revealed that dopaminergic neurons selectively failed to execute mitophagy. This method will be useful in providing a reliable readout of mitophagy in a range of experimental biomedical settings. —SMH

Cell **181**, 1176 (2020).

NEUROSCIENCE

"Flash and freeze" synaptic plasticity

The eponymous mossy fibers in the hippocampus of the brain connect the dentate gyrus to the CA3 pyramidal region and are implicated in short-term memories spanning a few seconds to minutes. How such memories are held in the circuit is a puzzle. One candidate is post-tetanic potentiation, a form of synaptic plasticity that decays within tens to hundreds of seconds and regulates synaptic strength at hippocampal mossy fiber–CA3 pyramidal neuron synapses. Vandael *et al.* simultaneously recorded from pairs of mossy fiber terminals and postsynaptic CA3 pyramidal neurons and followed measurements with "flash and freeze" electron microscopy. The authors observed that post-tetanic potentiation was induced by an increase in the size of the readily releasable pool of synaptic vesicles in granule cells. —PRS

Neuron 10.1016/j.neuron.2020.05.013 (2020).

QUANTUM INFORMATION

Set, exchange, measure, repeat

Successful protocols for quantum information processing and quantum computation depend on the reliable storage and manipulation of the quantum state of a qubit. Qubits, however, are prone to errors because complete isolation from the environment is not possible. Methods for correcting these errors must also contend with the fact that direct measurement of a qubit destroys it. Xue *et al.* describe a repetitive quantum nondemolition method on a two-qubit system in which the state of the main qubit is mapped onto a second qubit that acts as an ancilla. Repeated measurement of the ancilla qubit allows the main qubit to be maintained and read out with higher fidelity. Such a protocol provides a route to robust quantum information processing in solid-state systems. —ISO

Phys. Rev. X **10**, 021006 (2020).

SCIENCE AND POLICY

Testing peer review in government

To improve the role of science in government, some have called for more peer review. To test this approach, Ho and Larrimore Ouellette randomly assigned pending U.S. patent applications to external academic technical experts whose reviews were then provided to government patent examiners. Compared with matched, non–peer-reviewed applications, peer review increased examiners' efforts to search for earlier publications ("prior art"), increased citations to nonpatent literature, and reduced the propensity to initially grant the application. However, the effort was costly because of strict requirements (for example, time spent translating legalese) for how reviews could be introduced into the examination process. —BW

J. Empir. Leg. Stud. **17**, 190 (2020).

PHOTO: DAVID L. MOORE - AK/ALAMY STOCK PHOTO

Published by AAAS

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

FOREST AND CLIMATE

Risks to mitigation potential of forests

Much recent attention has focused on the potential of trees and forests to mitigate ongoing climate change by acting as sinks for carbon. Anderegg *et al*. review the growing evidence that forests' climate mitigation potential is increasingly at risk from a range of adversities that limit forest growth and health. These include physical factors such as drought and fire and biotic factors, including the depredations of insect herbivores and fungal pathogens. Full assessment and quantification of these risks, which themselves are influenced by climate, is key to achieving science-based policy outcomes for effective land and forest management. —AMS

Science, this issue p. 1327

EVOLUTION

The origins of pollination

Which came first, the flower or pollinating insects? Fossil evidence and molecular analyses using chloroplast DNA sequencing give conflicting estimates about when flowering plants first appeared, which has implications for how pollinating insects arose. In a Perspective, van der Kooi and Ollerton discuss the data and the possible explanations for the evolution of these interdependent species. —GKA

Science, this issue p. 1306

SOLAR CELLS

Perovskite decomposition in detail

Solar cells are subject to heating when operating in sunlight, and the organic components of hybrid perovskite solar cells, especially the commonly used methylammonium cation, can undergo thermal decomposition. Encapsulation can limit decomposition by bringing such reactions to equilibrium and can prevent exposure to damaging ambient moisture. Shi *et al*. examined several encapsulation schemes for perovskite films and devices by probing volatile products with gas chromatography–mass spectrometry (see the Perspective by Juarez-Perez and Haro). Pressure-tight polymer/glass stack encapsulation was effective in suppressing gas transfer and allowed solar cells containing methylammonium to pass harsh moisture and thermal cycling tests. —PDS

Science, this issue p. 1328; see also p. 1309

NEUROSCIENCE

Ensemble activity and perception

The mechanisms by which sensory percepts are encoded in neural ensembles are still incompletely understood. Chong *et al*. used single-spot optogenetic stimulation to control neuronal activity in mouse olfactory glomeruli in space and time. Animals were trained to recognize a learned activity pattern that was likely perceived as a specific odor. The authors then systematically varied the activity patterns by changing either the activated glomeruli or the timing between activation of glomeruli to evaluate their impact on odor recognition. Glomeruli that were activated early during the synthetic odor contributed more to odor recognition than glomeruli that were subsequently activated. This approach allows neuroscience to explain how features combine in complex patterns to generate perception. —PRS

Science, this issue p. 1329

NEUROSCIENCE

Secrets of Meissner corpuscles

The Meissner corpuscle, a mechanosensory end organ, was discovered more than 165 years ago and has since been found in the glabrous skin of all mammals, including that on human fingertips. Although prominently featured in textbooks, the function of the Meissner corpuscle is unknown. Neubarth *et al*. generated adult mice without Meissner corpuscles and used them to show that these corpuscles alone mediate behavioral responses to, and perception of, gentle forces (see the Perspective by Marshall and Patapoutian). Each Meissner corpuscle is innervated by two molecularly distinct, yet physiologically similar, mechanosensory neurons. These two neuronal subtypes are developmentally interdependent and their endings are intertwined within the corpuscle. Both Meissner mechanosensory neuron subtypes are homotypically tiled, ensuring uniform and complete coverage of the skin, yet their receptive fields are overlapping and offset with respect to each other. —PRS

Science, this issue p. 1330; see also p. 1311

CORONAVIRUS

Mobile symptom tracking

The rapidity with which severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spreads through a population is defying attempts at tracking it, and quantitative polymerase chain reaction testing so far has been too slow for real-time epidemiology. Taking advantage of existing longitudinal health care and research patient cohorts, Drew *et al*. pushed software updates to participants to encourage reporting of potential coronavirus disease 2019 (COVID-19) symptoms. The authors recruited about 2 million users (including health care workers) to the COVID Symptom Study (previously known as the COVID Symptom Tracker) from across the United Kingdom and the United States. The prevalence of combinations of symptoms (three or more), including fatigue and cough, followed by diarrhea, fever, and/or anosmia, was predictive of a positive test verification for SARS-CoV-2. As exemplified by data from Wales, United Kingdom, mathematical modeling predicted geographical hotspots of incidence 5 to 7 days in advance of official public health reports. —CA

Science, this issue p. 1362

MEASLES

Older origins of measles virus

Animal domestication by humans is thought to have given many pathogens an opportunity to invade a new host, and measles is one example of this. However, there is controversy about when measles emerged in humans, because the historical descriptions of measles are relatively recent (late ninth century CE). The controversy has persisted in part because ancient RNA is thought to be a poor target for molecular clock techniques. Düx *et al*. have overcome the ancient RNA challenge by sequencing a measles virus genome obtained from a museum specimen of the lungs of child who died in 1912 (see the Perspective by Ho and Duchêne). The authors used these and other more recent sequencing data in a Bayesian molecular clock–modeling technique, which showed that measles virus diverged from rinderpest virus in the sixth century BCE, indicating an early origin for measles possibly associated with the beginnings of urbanization. —CA

Science, this issue p. 1367; see also p. 1310

Published by AAAS

CHEMICAL ECOLOGY

A plant-herbivore information "arms race"

The consumption of plants by herbivores has driven the evolution of many diverse plant defense chemicals to which herbivores have constantly adapted. The transmission of chemical information at the community level is less known but important given the plethora of plant and herbivore species, especially in tropical communities. Zu *et al*. propose an information "arms race" approach to explain plant-herbivore chemical communication at the community level (see the Perspective by Solé). To test their conceptual framework, they used field data of herbivore-plant interactions and plant–volatile organic compound associations in a tropical dry forest. Their approach provides an understanding of the functioning and persistence of systems where individuals send and receive information in the form of signals to which other individuals react and, in turn, affect the behavior of other participants in these systems. —AMS

Science, this issue p. 1377; see also p. 1315

IMMUNOMETABOLISM

Inflammaging? Blame T cells!

Mitochondrial dysfunction in various tissues is a prominent characteristic of age-related deterioration, but it is unclear how mitochondrial dysfunction in particular cell types contributes to this process. Desdín-Micó *et al*. generated mice with T cells that were specifically deficient in a mitochondrial DNA–stabilizing protein. These animals exhibited multiple features associated with aging, including neurological, metabolic, muscular, and cardiovascular impairments. The defective T cells initiated an inflammatory program similar to that observed in older animals, a process called "inflammaging." Blocking the cytokine tumor necrosis factor–α or administering precursors of the cofactor nicotinamide adenine dinucleotide restored many of these symptoms of senescence. These findings may potentially inform future therapies for age-associated diseases, as well as cachexia and cytokine-release syndrome. —STS

Science, this issue p. 1371

PHASE SEPARATION

Drug partitioning in nuclear condensates

There is increasing interest in the function of phase-separated biomolecular condensates in cells because of their distinct properties and expanding roles in important biological processes. Klein *et al*. considered the fate of small-molecule therapeutics in the context of nuclear condensates (see the Perspective by Viny and Levine). They show that certain antineoplastic drugs have physicochemical properties that cause them to concentrate preferentially in condensates, both in vitro and in cancer cells. This property influences drug activity, and protein mutations that alter condensate formation can lead to drug resistance. Optimizing condensate partitioning may be valuable in developing improved therapeutics. —VV

Science, this issue p. 1386; see also p. 1314

VACCINES

Tempting T cells

Strategies to induce T cell responses during vaccination are difficult to execute. Esposito *et al*. tested a vaccine that uses the major histocompatibility complex class II invariant chain (Ii), which is important for antigen presentation to T cells, as an adjuvant in healthy volunteers. The so-called "prime-boost" strategy involved viral vectors and nonstructural antigens from hepatitis C virus. Inclusion of Ii boosted the magnitude and breadth of the T cell response. Work in a mouse vaccination model demonstrated that the Ii adjuvant enhanced proteasomal degradation of the vaccine antigens. This promising platform could be used to "tempt" T cells into vaccine responses, potentially resulting in successful vaccines for diseases that cannot be treated with conventional, antibody-driven vaccine protection. —LP

Sci. Transl. Med. **12**, eaaz7715 (2020).

ALLERGY

Finding contributors to allergy control

House dust mites (HDMs) are a major source of aeroallergens that trigger human allergic responses, including asthma. To characterize the heterogeneity of human $CD4^+$ T helper (T_H) cell responses to HDM antigens, Seumois *et al*. stimulated blood T cells with HDM peptides and isolated the freshly activated T cells for single-cell RNA sequencing. One T_H cell subset was more frequent in individuals without HDM allergy. These cells may function to restrain allergic T helper 2 (T_H2) cell responses. Individuals with both HDM allergy and asthma were enriched for interleukin-9–producing T_H2 cells. These findings provide a basis for designing more precise approaches to thwart pathogenic T_H cell responses contributing to allergic asthma. —IRW

Sci. Immunol. **5**, eaba6087 (2020).

Published by AAAS

REVIEW SUMMARY

FOREST AND CLIMATE

Climate-driven risks to the climate mitigation potential of forests

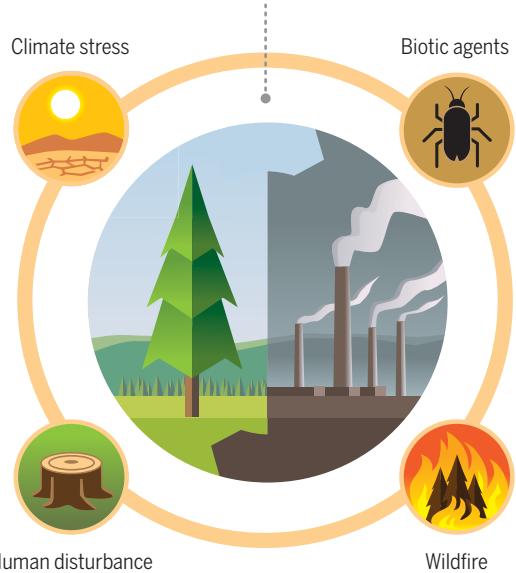
William R. L. Anderegg*, Anna T. Trugman, Grayson Badgley, Christa M. Anderson, Ann Bartuska, Philippe Ciais, Danny Cullenward, Christopher B. Field, Jeremy Freeman, Scott J. Goetz, Jeffrey A. Hicke, Deborah Huntzinger, Robert B. Jackson, John Nickerson, Stephen Pacala, James T. Randerson

BACKGROUND: Forests have considerable potential to help mitigate human-caused climate change and provide society with a broad range of cobenefits. Local, national, and international efforts have developed policies and economic incentives to protect and enhance forest carbon sinks—ranging from the Bonn Challenge to restore deforested areas to the development of forest carbon offset projects around the world. However, these policies do not always account for important ecological and climate-related risks and limits to forest stability (i.e., permanence). Widespread climate-induced forest die-off has been observed in forests globally and creates a dangerous carbon cycle feedback, both by releasing large amounts of carbon stored in forest ecosystems to the atmosphere and by reducing the size of the future forest carbon sink. Climate-driven risks may fundamentally compromise forest carbon stocks and sinks in the 21st century. Understanding and quantifying climate-driven risks to forest stability are crucial components needed to forecast the integrity of forest carbon sinks and the extent to which they can contribute toward the Paris Agreement goal to limit warming well below 2°C. Thus, rigorous scientific assessment of the risks and limitations to widespread deployment of forests as natural climate solutions is urgently needed.

ADVANCES: Many forest-based natural climate solutions do not yet rely on the best available scientific information and ecological tools to assess the risks to forest stability from climate-driven forest dieback caused by fire, drought, biotic agents, and other disturbances. Crucially, many of these permanence risks are projected to increase in the 21st century because of climate change, and thus estimates based on historical data will underestimate the true risks that forests face. Forest climate policy needs to fully account for the permanence risks because they could fundamentally undermine the effectiveness of forest-based climate solutions.

Here, we synthesize current scientific understanding of the climate-driven risks to forests and highlight key issues for maximizing the effectiveness of forests as natural climate solutions. We lay out a roadmap for quantifying current and forecasting future risks to forest stability using recent advances in vegetation physiology, disturbance ecology, mechanistic vegetation modeling, large-scale ecological observation networks, and remote sensing. Finally, we review current efforts to use forests as natural climate solutions and discuss how these programs and policies presently consider and could more fully embrace physiological, climatic, and permanence uncertainty about the future of forest carbon stores and the terrestrial carbon sink.

Forests as natural climate solutions face fundamental limits and underappreciated risks



Effective use of forests as natural climate solutions depends on accounting for climate-driven risks, such as fire and drought. Leveraging cutting-edge scientific tools holds great promise for improving and guiding the use of forests as natural climate solutions, both in estimating the potential of carbon storage and in estimating the risks to forest carbon storage.

OUTLOOK: The scientific community agrees that forests can contribute to global efforts to mitigate human-caused climate change. The community also recognizes that using forests as natural climate solutions must not distract from rapid reductions in emissions from fossil fuel combustion. Furthermore, responsibly using forests as natural climate solutions requires rigorous quantification of risks to forest stability, forests' carbon storage potential, cobenefits for species conservation and ecosystem services, and full climate feedbacks from albedo and other effects. Combining long-term satellite records with forest plot data can provide rigorous, spatially explicit estimates of climate change–driven stresses and disturbances that decrease productivity and increase mortality. Current vegetation models also hold substantial promise to quantify forest risks and inform forest management and policies, which currently rely predominantly on historical data.

A more-holistic understanding and quantification of risks to forest stability will help policy-makers effectively use forests as natural climate solutions. Scientific advances have increased our ability to characterize risks associated with a number of biotic and abiotic factors, including risks associated with fire, drought, and biotic agent outbreaks. While the models that are used to predict disturbance risks of these types represent the cutting edge in ecology and Earth system science to date, relatively little infrastructure and few tools have been developed to interface between scientists and foresters, land managers, and policy-makers to ensure that science-based risks and opportunities are fully accounted for in policy and management contexts. To enable effective policy and management decisions, these tools must be openly accessible, transparent, modular, applicable across scales, and usable by a wide range of stakeholders. Strengthening this science-policy link is a critical next step in moving forward with leveraging forests in climate change mitigation efforts. ■

ON OUR WEBSITE

Read the full article at https://dx.doi.org/10.1126/science.aaz7005

The list of author affiliations is available in the full article online.
*Corresponding author. Email: anderegg@utah.edu
Cite this article as W. R. L. Anderegg *et al.*, *Science* **368**, eaaz7005 (2020). DOI: 10.1126/science.aaz7005

ILLUSTRATION: DAVID MEIKLE

REVIEW

FOREST AND CLIMATE

Climate-driven risks to the climate mitigation potential of forests

William R. L. Anderegg[1]*, Anna T. Trugman[2], Grayson Badgley[1], Christa M. Anderson[3], Ann Bartuska[4], Philippe Ciais[5], Danny Cullenward[6], Christopher B. Field[7], Jeremy Freeman[8], Scott J. Goetz[9], Jeffrey A. Hicke[10], Deborah Huntzinger[11], Robert B. Jackson[7,12], John Nickerson[13], Stephen Pacala[14], James T. Randerson[15]

Forests have considerable potential to help mitigate human-caused climate change and provide society with many cobenefits. However, climate-driven risks may fundamentally compromise forest carbon sinks in the 21st century. Here, we synthesize the current understanding of climate-driven risks to forest stability from fire, drought, biotic agents, and other disturbances. We review how efforts to use forests as natural climate solutions presently consider and could more fully embrace current scientific knowledge to account for these climate-driven risks. Recent advances in vegetation physiology, disturbance ecology, mechanistic vegetation modeling, large-scale ecological observation networks, and remote sensing are improving current estimates and forecasts of the risks to forest stability. A more holistic understanding and quantification of such risks will help policy-makers and other stakeholders effectively use forests as natural climate solutions.

Terrestrial ecosystems currently absorb ~30% of human carbon emissions each year (*1*), and forests account for the vast majority of this uptake [an estimated 8.8 Pg CO_2e year^{-1} of a total land carbon uptake of 9.5 Pg CO_2e year^{-1} over 2000–2007, where CO_2e denotes CO_2 equivalents (*2*, *3*)]. A broad body of literature has focused for decades on the role of forests in the climate system (*4*–*6*), and forest-based natural climate solutions (F-NCSs) have experienced growing interest in recent years as a potentially major contributor to meeting Paris Agreement carbon targets (*7*–*10*). Forest-based strategies might provide up to 7 Pg CO_2e of climate mitigation per year by 2030 at a carbon price of $100 per Mg CO_2e, which is by far the largest potential category of natural climate solutions (NCSs) (*7*). Furthermore, many of these forest-based strategies are likely to have substantial cobenefits for biodiversity, ecosystem services, and conservation (*9*, *11*).

Carbon policy that includes F-NCSs is building around the world (Fig. 1). For example, California has recognized 133 Tg CO_2e in benefits from forest carbon offset projects in the United States between 2013 and 2019, with these credits making up a meaningful share of the compliance with the state's cap-and-trade program (*12*). National and subnational government policies to reduce emissions have included forest projects, with policies in Japan, Australia, New Zealand, and British Columbia, Canada (Fig. 1). Additionally, many F-NCS projects have occurred under the framework of the United Nations' Reducing Emissions from Deforestation and Degradation (REDD+) (*13*, *14*) and under local and national emissions reduction goals.

F-NCS projects include a wide array of project types but broadly fall into four categories: (i) avoided forest conversion (i.e., avoided deforestation); (ii) reforestation; (iii) improved management of natural forests; and (iv) improved forest plantation practices (*7*, *9*, *15*). An overarching commonality is that all F-NCS projects strive for permanence, the principle that forests store carbon removed from the atmosphere in plants and soils over time horizons of 50 to 100 years or longer. Given that a large fraction of human emissions of CO_2 remain in the atmosphere for centuries to millennia (*16*), the permanence of forest carbon on century time scales is essential for effective climate change mitigation.

Fundamental questions remain, however, about the fate of carbon stored in forests in a rapidly changing climate, particularly the extent to which climate change and climate-driven changes in disturbance regimes might compromise forest permanence (*17*–*19*). Climate-induced tree mortality events have been widely observed across the globe over the past few decades (*20*, *21*). In addition to direct climate impacts on trees like drought events, additional disturbance agents including wildfire and insect outbreaks are sensitive to climate and have major carbon cycle consequences for forests (*22*–*25*). The biomass dynamics of an estimated 44% of forests globally are strongly sensitive to stand-replacing disturbance (including harvest) (Fig. 2) (*26*). Further, climate-driven tree mortality and disturbances are nonstationary (they change with time) and are projected to increase with climate change (*25*). Finally, due in part to the large uncertainties about climate impacts, CO_2 fertilization, and disturbances in forests (*27*), Earth system model projections over the 21st century indicate that terrestrial ecosystems could sequester as much as 36.7 Pg CO_2e year^{-1} or release as much as 22 Pg CO_2e year^{-1} by 2100 for a high-emissions scenario (*28*).

Nonstationary risks from climate change have the potential to compromise the current land carbon sink, the success of F-NCS projects, and tree-based bioenergy projects, such as some types of bioenergy with carbon capture and sequestration (BECCS) (*17*, *27*, *29*–*31*). Nonstationary changes in disturbance rates or long-term shifts in ecosystems (e.g., loss of forest) are what fundamentally determine the permanence of forest carbon stocks at large scales. The net carbon cycle effects of stationary disturbance regimes at landscape scales are small because carbon emissions from recently disturbed areas in one part of the landscape are compensated by sinks in regrowing areas (*32*, *33*). However, forests are already facing substantial and increasing climate-driven risks that could fundamentally undermine their collective ability to take up and store carbon over the 21st century (*19*, *22*, *34*, *35*). Thus, nonstationary permanence risks must be rigorously assessed using the best available scientific tools and datasets and be included in policy and project planning.

Here, we provide a review of key climate-driven risks to forest carbon permanence (i.e., disturbances that could lead to substantial losses in forest carbon stocks) and how these risks are expected to change in the future. We assess key climate-driven risks from the following perspectives: (i) carbon cycle impacts; (ii) data on historical patterns and risk levels; (iii) current mechanistic understanding and modeling approaches; and (iv) projections of nonstationary risk for future climate change scenarios. We then discuss how ongoing and planned F-NCS projects and policies currently account for permanence risk. Next, we provide a roadmap for the rigorous assessment

[1]School of Biological Sciences, University of Utah, Salt Lake City, UT 84113, USA. [2]Department of Geography, University of California, Santa Barbara, Santa Barbara, CA 93106, USA. [3]World Wildlife Fund, Washington, DC 20037, USA. [4]Resources for the Future, Washington, DC 20036, USA. [5]Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre Simon Laplace CNRS CEA UVSQ Gif sur Yvette, 91191, France. [6]Stanford Law School, Stanford, CA 94305, USA. [7]Woods Institute for the Environment, Stanford University, Stanford, CA 94305, USA. [8]CarbonPlan, San Francisco, CA 94110, USA. [9]School of Informatics and Computing, Northern Arizona University, Flagstaff, AZ 86011, USA. [10]Department of Geography, University of Idaho, Moscow, ID 83844, USA. [11]School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ 86011, USA. [12]Department of Earth System Science and Precourt Institute, Stanford University, Stanford, CA 94305, USA. [13]Climate Action Reserve, Los Angeles, CA 90017, USA. [14]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08540, USA. [15]Department of Earth System Science, University of California Irvine, Irvine, CA 92697, USA.
*Corresponding author. Email: anderegg@utah.edu

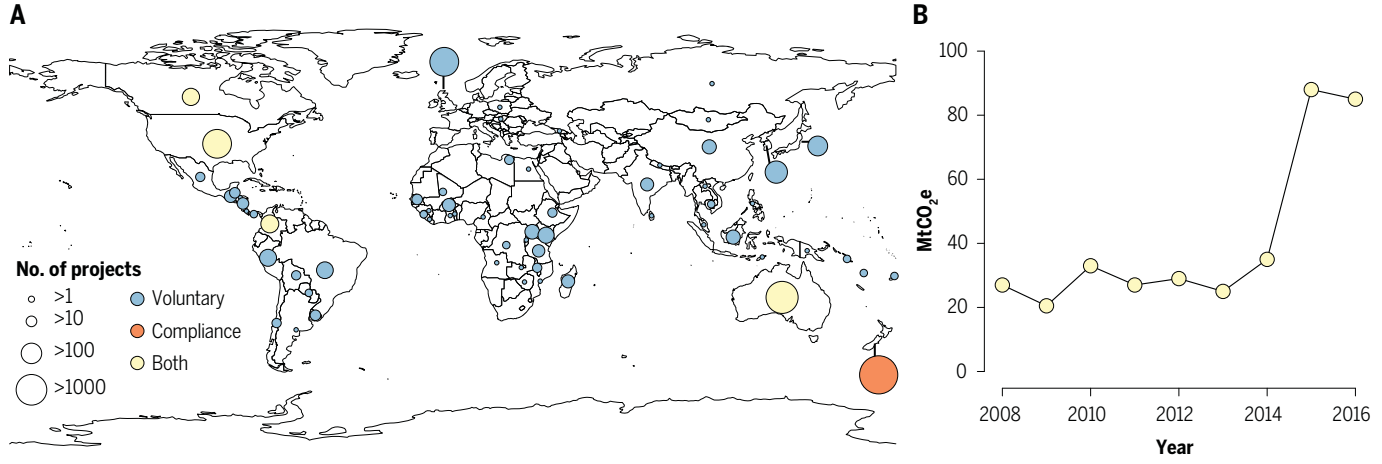


Fig. 1. F-NCSs are being used around the world. (**A**) F-NCS projects around the world by country and type of mechanism (voluntary, compliance, or both) as of January 2017, the most recent data available. (**B**) Carbon volume in million metric tons of carbon dioxide equivalent (Mt CO_2e) covered in compliance and voluntary forest carbon projects 2008–2016. Redrawn from (*137*).

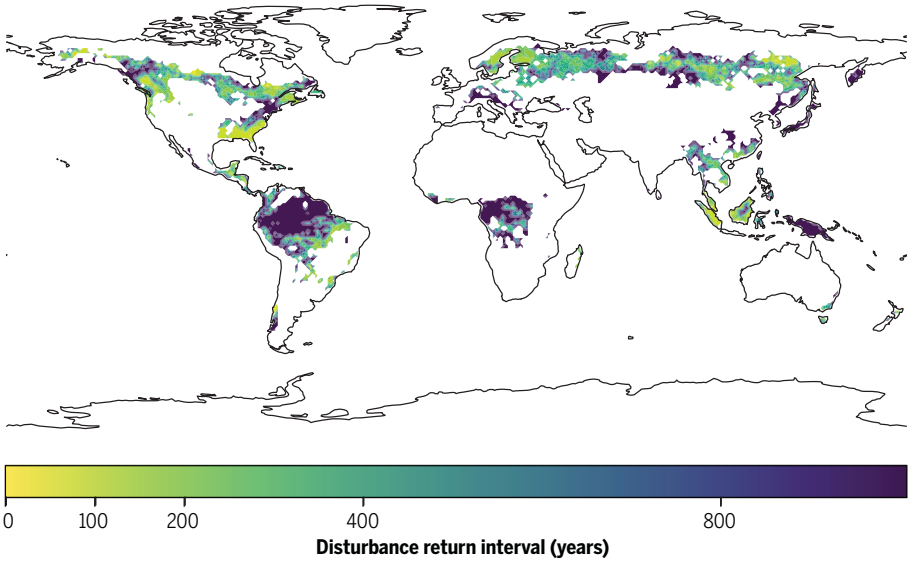


Fig. 2. Sensitivity of global forest biomass dynamics to stand-replacing disturbance (excluding human land use changes) captured by disturbance return interval (years). Warm colors indicate areas where biomass dynamics are highly sensitive to the frequency of stand-replacing disturbance and cool colors indicate areas that are relatively less sensitive. Redrawn from (*26*).

of policy-relevant permanence risks by combining a broad array of datasets and models. Finally, we discuss ways forward to bridge science-policy gaps so that the best available science can inform the future of forest carbon sinks and help promote the success of natural climate solutions.

Major climate-sensitive permanence risks

Fire

Fires in forests are perhaps the most well-quantified global disturbance and permanence risk. Between 1997 and 2016, an average of ~500 million ha of land burned each year, most of which is outside of forest ecosystems (*36*). Although burned area is declining in grasslands and savannas, burned area is increasing in many tropical, temperate, and boreal forest ecosystems (*36*). Fire in forests emits ~1.8 Pg CO_2e $year^{-1}$ (*37*, *38*). Fire accounts for ~12% of stand-replacing disturbances in forest ecosystems annually (*26*) and is particularly important in key forest regions like the western United States, Australia, Mediterranean-type climates, and boreal forests in North America and Asia (*39*, *40*). Climate-driven changes in fire regimes can affect permanence both through changes in burned area and through changes in fire behavior (i.e., fire temperature or scorch height) that influence tree mortality and—in many temperate and boreal forests—the amount of fuel consumption in organic surface soil layers (*41*). Multiple satellite datasets have mapped fire-burned area and emissions at moderate or high spatial resolution globally from the late 1990s to the present (*37*, *42*) and extend even further back at low resolution globally (*43*) and at high resolution in some regions, such as the United States (*44*), Australia (*45*), and Canada (*46*). Paleoclimate reconstructions of fire also hold promise for informing projections of future fire regimes (*44*). Thus, long-term fire data are widely available for assessing historical permanence risks.

A wide variety of fire models also exist, including both empirical and mechanistic models that differ in complexity and mechanisms considered (*44*). Empirical models (*47*, *48*) and mechanistic models (*49*, *50*) broadly project increases in fire activity and permanence risks with climate change but with substantial regional heterogeneity (*44*). Mechanistic fire models are an active area of research (*50*, *51*), and improved fire models are being incorporated into terrestrial biosphere models (*50*). These models aim to simulate the complex dynamics between vegetation, climate, and fire, as well as changes in human land use and populations that influence fire ignition and fire spread (see below).

Drought

Globally, droughts represent a major and widespread permanence risk, underscored by the explosion of research relating to drought-induced tree mortality that has been done in the past decade (*21*, *52*). Drought events have major impacts on forest carbon cycling through declines in productivity and carbon losses through mortality (*20*, *27*). Major climate extremes explain up to 78% of the variation in global gross primary productivity in the past

30 years, and severe droughts made up ~60 to 90% of the largest extremes (*53*). As an example, the severe 2011–2015 drought in California killed more than an estimated 140 million trees and drove the full carbon balance of the state's ecosystems to be a net source of −600 Tg CO_2e from 2001 to 2015, which is equivalent to ~10% of the state's greenhouse gas emissions over that period (*54*). A 2011 drought in Texas killed 9.5% of tree cover across the state, and much of the canopy loss occurred in areas that exceeded specific climatic thresholds (*55*). Increasingly severe drought in Australia has also led to systematic increases in tree mortality and composition changes (*56*).

Substantial historical data on drought risks are available from a variety of data sources, although such data are relatively less direct and less detailed than information on burned area at the global scale. Climate data from weather stations and reanalysis products allow for the calculation of meteorological and agricultural drought and aridity metrics (*57*). The utility of these datasets in estimating spatial patterns of permanence risk depends on the extent to which forest vulnerability (e.g., mortality) correlates with drought severity and/or average aridity, which has been observed in several meta-analyses (*58*–*60*). Remote sensing data can provide spatial patterns of drought impacts on productivity and, in some cases, drought-driven mortality (*61*, *62*). The central difference compared with fire is that drought-driven tree mortality is often more widespread (i.e., occurring over large regions because of the widespread nature of drought) and more diffuse (i.e., a smaller number of trees are killed per area). Furthermore, mortality can occur both during and for multiple years after a drought event, which can lead to underestimation of the impact of drought on forest carbon (*63*). As a result of the more diffuse signal in space and time, drought-driven mortality can be more challenging to detect with moderate-resolution satellite imagery, which complicates the quantification of carbon cycle impacts of drought on tree mortality (*21*).

Drawing on a broad set of tree ecophysiological research, mechanistic vegetation models have rapidly improved in simulating drought risks to forest permanence. The latest results from these models suggest that regional and global drought risk estimates will be possible in the short term (*21*). Drought is thought to drive tree mortality primarily through a failure of the plant water transport (hydraulic) system (*64*–*66*), and the biophysical processes that mediate this failure have been relatively well understood for decades (*67*). Species' hydraulic traits have been shown to explain patterns in mortality risk within ecological communities (*58*, *59*), which provides a promising avenue for using widely measured plant functional traits in permanence risk assessments. Furthermore, vegetation and land surface models have recently incorporated key aspects of hydraulic transport (see below), which should enable drought-related permanence risk forecasting in the near future.

Biotic agents

Biotic disturbance agents, including insects and pathogens, cause substantial tree mortality globally. For example, bark beetles, which feed on tree phloem and introduce fungi that interrupt tree water transport, have killed billions of trees across millions of hectares of land in temperate and boreal coniferous forests in the past two decades (*68*–*70*) and have converted large regions of the Canadian boreal forest from a sink to a source over the course of a decade (*34*). Defoliators feed on leaves and can kill trees after multiple years of severe damage. Widespread tree mortality has occurred from defoliators in both coniferous and broad-leaved forests in temperate and boreal regions (*24*, *71*). In addition to these native biotic agents, non-native invasive biotic disturbance agents are responsible for killing many trees globally. Prominent examples include *Phytophthora*-induced sudden oak death and the emerald ash borer in the United States and the red turpentine beetle in China (*72*).

Aerial surveys and remote sensing imagery have provided estimates of and constraints on permanence risks from some biotic agents over the past 40 years (*40*). The spatial patterns and impacts of major biotic agent outbreaks have been extensively mapped from surveys, plot measurements, and agency reports in many regions (*70*). For example, in the western United States, bark beetle–caused tree mortality during 1997–2010 affected 5% of aboveground tree carbon stocks, which is about the same as was affected by wildfires during that period (*40*). As with drought, tree mortality from biotic agents is often diffuse, which creates a challenge for change detection using satellite imagery. Detection efficiency depends on the severity of the outbreak, instrumental spatial and spectral resolution, the quality of ground observations, the duration and frequency of the satellite observations, and the underlying change-detection algorithms. Attribution of tree mortality to a specific biotic agent—and separation from drought influences—is difficult; however, high confidence in attribution may not be needed for an initial assessment of permanence risk patterns.

Climate affects the outbreaks of many biotic disturbance agents. In the case of bark beetles, warmer conditions increase winter survival and increase life stage–development rates (*73*). Drought stresses host trees, thereby decreasing defenses, altering foliage quality, and leaving trees more susceptible to attack (*74*). Ambient moisture can influence pathogen survival and spread (*75*, *76*). However, insect-host systems are complicated by a variety of factors and interactions, and our understanding about the relative importance and functional relationships of climate drivers is limited (*77*). Thus, tree mortality from biotic disturbance agents remains remarkably difficult to model and predict, especially over larger spatial scales and longer time periods (*78*, *79*). Although challenging, the vulnerability of a forest stand can be partially estimated with models that evaluate stand structure (species composition, stem density, size, age, vigor, etc.) for many biotic disturbance agents (*80*). Biotic agent sensitivity to different climate influences (e.g., winter mortality or drought stress) have been assessed for prominent insect species, for example in the western United States (*81*).

The permanence risks from biotic agents to forests are likely nonstationary and are expected to increase substantially in the future (*25*). Integrated biotic agent models that combine multiple drivers, including key climate sensitivities, and predict tree mortality are needed to assess the permanence of forest carbon (*79*). These models are challenging to develop even for the best-understood agents and are limited in number (*82*). Thus, predictive tools for biotic agent disturbance are the most limited among the disturbance types we cover here because of the following: (i) the diversity of insects and pathogens across forest ecosystems; (ii) the introduction of nonnative biotic agents, which often cannot be predicted; and (iii) the complicated cross-scale dynamics between climate, biotic agent populations, and tree populations (*77*, *79*).

Other disturbances

Other disturbances—particularly storms and wind-driven events, snow and ice events, and lightning—can also influence forest ecosystem carbon cycling (*25*, *83*). These disturbance events can matter for local- to regional-scale carbon cycling in some areas but are thought to have relatively minor-to-modest global effects (*25*, *53*). Hurricanes damage coastal forests and can have pronounced impacts on carbon budgets. For example, Hurricane Katrina damaged 320 million large trees that contained 385 Tg CO_2e (*83*), and tropical cyclones had a net effect of a modest carbon source in the 20th century across U.S. forests (*33*). However, a key question for wind and other disturbances is whether projected future trends indicate that risks are likely to increase. A recent meta-analysis identified some projected increases in wind disturbance in some regions, but it identified little directional change in other regions and little projected change in other disturbance events such as snow and ice events (*25*).

In addition to large-scale disturbances, climate-driven shifts in the ranges of tree species and forest community assemblages are already occurring and are likely to be even more

widespread in the future (*84*–*86*). Changes in community composition can have major carbon cycle impacts (*85*, *87*, *88*) and can mediate forest vulnerability to disturbance (*89*, *90*). A broad body of literature has explored potential range shifts and changes in community composition, although major uncertainties remain in future projections (*84*). Critically, adequate tree regeneration rates are important for maintaining forest permanence over the long term; thus, improved understanding and models of the climate and nonclimate (e.g., non-native species) drivers of regeneration rates are needed (*91*, *92*).

Disturbance interactions

Disturbances that can drive risks to forest permanence often co-occur or interact at multiple spatial and temporal scales (*25*, *93*). For example, fire often co-occurs with drought in many regions globally (*94*). Drought and biotic agent attacks also often co-occur and can interact in complex ways to mediate tree mortality (*79*, *95*, *96*). Although interactions among disturbances can either dampen or magnify carbon cycle impacts on forests, a recent meta-analysis found that the interaction effects typically magnify carbon losses for most of the climate-sensitive disturbances and regions (*25*).

Human interactions

Human actions can increase or decrease climate-related permanence risks. Human appropriation of forest biomass is ~9.5 Pg CO_2e $year^{-1}$ (*97*), which is more than annual fire emissions. Thus, socioeconomic changes that alter human interactions with forest biomass may have large consequences for permanence of forests. Thorough treatment of these interactions has been covered elsewhere (*20*, *49*, *98*), and we briefly highlight a few key interactions. In particular, land management such as forest thinning and fuels reduction can decrease the risks of fire, drought, and attacks by biotic agents in some forests (*99*). Humans are, however, key ignition sources for fires around the world (*36*), so changes in human populations, land-use, policy, and behavior can affect projections of fire risk (*49*). Humans also frequently fight or suppress fires, and in some cases biotic agent activity, to minimize and mitigate negative outcomes on livelihoods, ecosystem services, and carbon cycling. We note that human management actions, such as forest harvest and sustainable forest management, can be important mechanisms to maintain forest carbon sinks and could be used strategically to decrease permanence risks. Finally, the introduction of non-native invasive insects and pathogens by humans over the past few centuries has led to substantial tree mortality that continues today (*100*), and further introductions in the future could have similar effects.

Current efforts to address permanence risk

The degree to which current F-NCS efforts include permanence risks varies enormously, and very few projects to date have considered nonstationary risks from climate change. Some F-NCS projects have no explicit way to address permanence risks. Other F-NCS efforts include at least some estimate of permanence risks and contain mechanisms such as a buffer pool that can account for risks across a portfolio of forest projects. Even in these cases, however, the data underpinning many protocols' risk assessments are often unclear and, where delineated, are based on average historical conditions with little spatial or ecosystem-specific granularity. Therefore, additional consideration should be given to the following: (i) whether risks have been adequately assessed and (ii) whether nonstationary risks due to climate change are likely. Additionally, spatially explicit and regularly updated risk data would enable a quantitative risk assessment of given portfolios and inform project planning.

Crucially, risk estimates developed from historical data are highly unlikely to be adequate in capturing the increasing permanence risks of many disturbances—particularly those of fire, drought, and biotic agents—over the full 21st century (Fig. 3). For example, the mean 100-year integrated risk of moderate and severe wildfire across all U.S. ecosystems has already approximately doubled from ~4% over 1984–2000 to ~8% in 2001–2017 (Fig. 4), and much of this shift can be directly attributed to climate change (*101*). Furthermore, increases in the spatial extent and frequency of fire, drought, and biotic disturbances are expected in the future from climate change, yet relevant forests may not have experienced these disturbances in the recent past, which suggests that historical data may not capture future risk. We discuss several prominent examples of different approaches currently used to account for permanence risks below.

California forest offset program example

Pursuant to the Global Warming Solutions Act (A.B. 32), California established a cap-and-trade program that includes a forest offset program. This program is one of the largest compliance offset programs in existence and thus is an important case study (*15*). The offset program defines forest project permanence on a 100-year basis and deals with risk of unintentional loss of carbon stocks by using a buffer pool approach (*15*, *102*, *103*). Forest offset project owners are required to contribute a percentage of forest carbon credits earned from their projects to a common buffer pool account. Buffer pool credits are retired (i.e., removed) to mitigate for any unintentional carbon loss. The buffer pool is capitalized by taking a share of project credits (indicated in parentheses) for the following risks: wildfire (2 to 4%), disease or insects (3%), other natural catastrophes (e.g., drought, hurricane, tornado, or wind) (3%), overharvesting (0 to 2%), conversion to a nonforest land use (0 to 2%), and bankruptcy (1 to 5%) (*103*). As of 2019, an average of 16% of credits earned was submitted to the buffer pool in recognition of these risks

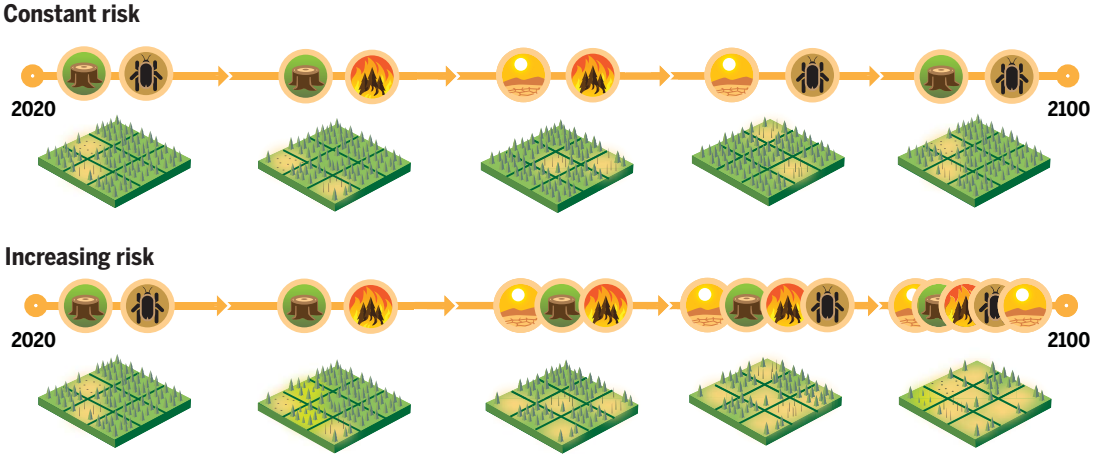


Fig. 3. Increasing climate-driven disturbance risk over time has major impacts on forest carbon. Conceptual diagram of stationary or constant (**Top**) versus nonstationary or increasing (**Bottom**) permanence risks from disturbance at a landscape scale in a changing climate. Disturbance events are illustrated in the circles and include fire, drought, biotic agents, and human disturbance.

ILLUSTRATION: DAVID MEIKLE

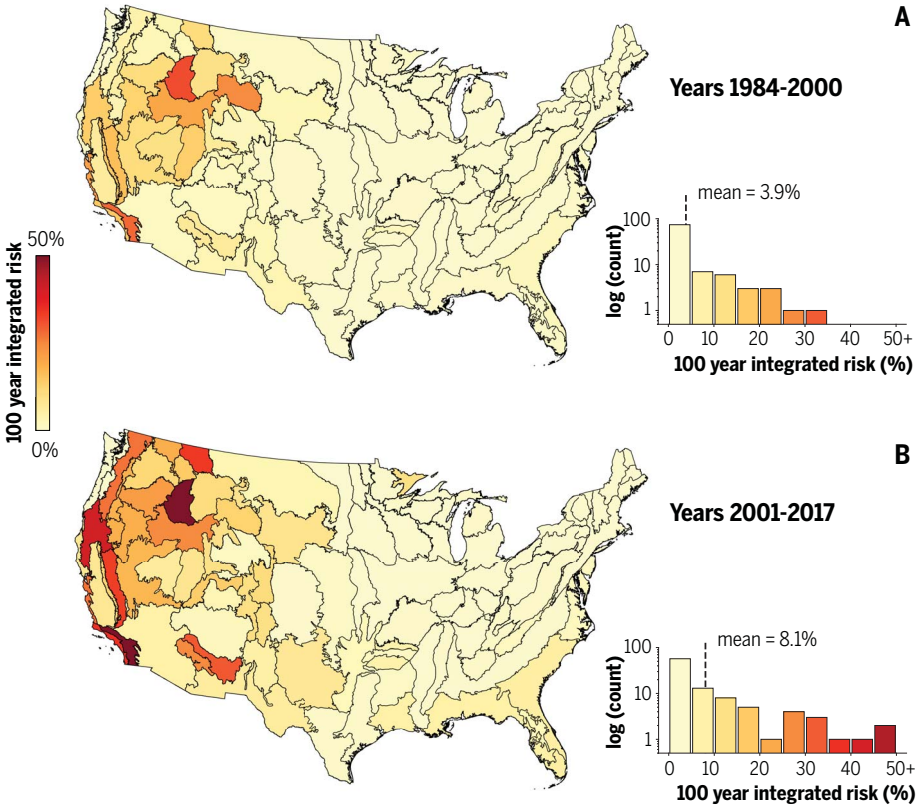


Fig. 4. Climate change has already increased fire risk in ecosystems. (**A** and **B**) Integrated 100-year fire risk of moderate- or high-severity fire from the Monitoring Trends in Burn Severity (MTBS) dataset based on fire occurrences in years 1984–2000 aggregated to ecoregions (A) and for fire occurrences in years 2001–2017 aggregated to ecoregions (B). Fire risk was computed as follows. First, within each ecoregion and year, a pixel-wise burn probability was computed as the fraction of pixels in that ecoregion labeled as moderate or severe fire, and these probabilities were then averaged in each time period. To project an integrated 100-year risk, we computed the probability of any pixel experiencing at least one fire under a binomial distribution with 100 trials and success probability given by the pixel-wise annual risk described above. This is a simple analysis that does not account for spatial or temporal autocorrelation or attempt to model any drivers of fire risk. Raw data obtained from www.mtbs.gov/direct-download, and Python code to create figures is available at https://github.com/carbonplan/forest-climate-risks.

to permanence (*104*). Intentional loss of forest carbon must be compensated by the project owner, and unintentional losses are absorbed by the buffer pool, which plays the role of a carbon insurance system. Any unintentional reversal can draw on the buffer pool, no matter the type.

Given the risk percentages above, California's forest offset project portfolio currently uses an 8 to 10% buffer for the climate-sensitive permanence risks discussed here. For fire risk, there are two levels of risk assessed: 2% for projects that have conducted fire risk–reduction work and 4% for projects that have not conducted fire risk–reduction work (*103*). These risk levels are applied at a constant level across the entire United States and thus do not account for ecoregion-level or spatial differences in historical permanence risks (Fig. 4). Crucially, none of the risk categories explicitly accounts for climate change. Thus, a central question moving forward is how the best available science can inform risk estimates to reflect the combination of current and projected nonstationary risks over >100 years.

Other approaches

Several major offset organizations in the United States and other jurisdictions have used a similar buffer pool approach to manage permanence risks. Under Japan's Certification Standard for Forest Carbon Sink, 3% of credits, total, are allocated to a buffer pool (*105*). The New Zealand Emissions Trading Scheme takes an approach that is different from the buffer pool and instead addresses permanence risk by instituting two types of offset credits: temporary and permanent (*98*). Although a buffer pool approach remains the primary method of addressing permanence risks, other insurance approaches, such as pooling risk across a wide array of carbon removal projects (including nonforest projects), have been proposed (*106*).

These are all approaches taken by forest offset programs, where emphasis is placed on measuring exact tons of carbon sequestered by forests to offset fossil fuel emissions. Other kinds of F-NCS projects, where forests are not being used as offsets but instead strive to contribute to mitigation more broadly—such as results-based finance projects (e.g., the California Climate Investments initiative)—have not consistently implemented methods for explicitly assessing permanence risk to the same degree that offset projects have.

A roadmap for rigorous permanence assessment

Rigorous quantification of current and future permanence risk is increasingly possible using vegetation ecophysiology, disturbance ecology, mechanistic vegetation modeling tools, large-scale ecological observation networks, and remote sensing imagery and products (Fig. 5). Leveraging these rapidly advancing models and observations should enable the estimation of permanence risk at continental and global scales. Furthermore, both empirical and mechanistic models can be driven by historical data and projections of climate, land use, and land management, given scenarios of human decisions. If possible, such new risk estimates should be spatially explicit to support F-NCS project planning. We discuss some of the key tools and datasets below, along with how they can be productively integrated. Integrating these diverse datasets and tools is urgent but often challenging because of the wide range of temporal and spatial scales. An additional key scientific challenge is the better understanding and testing of the effectiveness of human interventions (e.g., forest management) in decreasing risks to permanence across forest systems.

Forest plot and inventory data

High-quality ground-based data, such as those provided by permanent forest inventory networks, play a key role in rigorous permanence risk assessment. Given the patchy and dispersed nature of drought- and biotic agent–driven disturbances, and the need to evaluate remote sensing observations, successful integration of field data with high–spatial resolution remote sensing data will be essential for deriving global permanence risk maps and testing mechanistic models (*21*, *26*). Many countries have well-established inventory networks where both tree growth and mortality are measured at low temporal frequency, typically every 5 to 10 years. Scientific plot networks such as the RAINFOR, AFRITRON (*107*), and Forests-GEO networks (*108*) will be helpful in other regions where inventories do not currently exist or are not available. Spatial

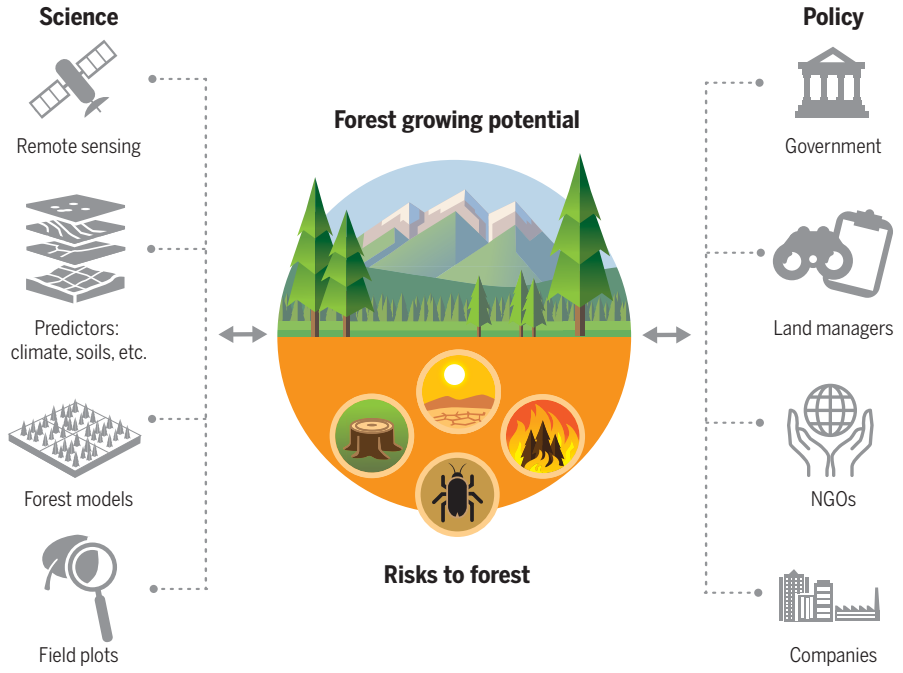


Fig. 5. Bridging science-policy divide on F-NCS projects. Key information needed at the science-policy nexus includes the carbon storage (current and potential) and the risks to forest permanence, among others. Central components of a rigorous scientific quantification of this information are presented on the left, and example key stakeholder groups [such as nongovernmental organizations (NGOs)] are presented on the right.

coverage, data availability, ease of access and use, and scale mismatches between remote sensing and ground plots remain important barriers to the widespread leveraging of these datasets.

Mechanistic models

Mechanistic vegetation models are one critical tool used for understanding how changes in climate could drive changes in ecosystem composition, structure and function, and allow projections of nonstationary forest permanence risks with novel climate regimes. Vegetation models simulate water, energy, and carbon fluxes and—when coupled to atmospheric models—ecosystem feedbacks to climate. Several critical advances in the representation of ecosystem ecology and ecophysiology make mechanistic vegetation models well suited for understanding forest permanence risk with climate change. However, some unresolved challenges remain, and considerable diversity exists in how mortality and the types of disturbance events are represented across models (*109*). Here, we summarize key model capabilities and current challenges as they relate to permanence risk prediction.

Recent advances in representations of ecosystem heterogeneity in mechanistic vegetation models hold strong promise for capturing forest permanence risks to drought, fire, and insect disturbance. Demographic ecosystem and stochastic gap models, two types of mechanistic vegetation models, capture ecosystem heterogeneity by resolving trees by size class, density, and plant functional type (*110*). Stochastic gap routines within demographic models represent disturbance-driven changes in ecosystem structure caused by processes such as treefall and fire, and they resolve the subsequent changes in local microenvironment caused by changes in canopy structure, light availability, and plant water demand (*111*). Further, some demographic ecosystem models include dynamic vegetation and simulate shifts in vegetation functional types, forest structure, and composition change in response to climate (*112*).

Although forest demographic processes are important for forecasting the permanence of forest carbon, representations of forest demography remain highly uncertain and are not widely included in large-scale vegetation models (*110*, *113*, *114*). Although the explicit representation of fire-, drought-, and insect-driven mortality is lacking in most models, the physiological and ecological processes important for capturing these types of mortality events have been incorporated in some mechanistic vegetation models. Further, an improved representation of mortality processes is a high priority in the vegetation modeling community (*26*, *113*–*115*). Thus, there is growing potential for mechanistic vegetation models to predict nonstationary permanence risks in response to disturbance. For example, tree size and density affect fire mortality risk, the potential for insect attack, and drought-driven mortality potential. Additionally, the availability of simple and predictive representations of vegetation water transport (i.e., hydraulic) processes makes it relatively straightforward to include process-based drought recovery and mortality even in large-scale vegetation models (*116*–*118*). Climate-driven disturbances have been implemented in vegetation models without explicit demography as well—for example, fire in the Community Land Model (*119*).

Modeling and predicting biotic agent–driven mortality, however, remains one of the largest unresolved challenges discussed in this review. Currently, biotic agent–driven mortality is often included in vegetation models as an implicit process included in density-independent, background mortality rates. This representation is particularly problematic because biotic-driven mortality is highly heterogeneous, affects different physiological processes depending on the type of insect or pathogen, is often responsive to climate, and can lead to catastrophic mortality events (*34*, *95*, *96*, *120*). Thus far, mechanistic vegetation models have been most useful in assessing the carbon cycle implications of insect disturbance rather than actually predicting insect-driven mortality events (*120*).

As a result of the above factors, large diversity exists in model representations of different mortality processes, scales, and structures. Some features of disturbance are also inherently hard to predict, such as the timing, extent, and magnitude of events (*121*). Because of model diversity as well as uncertainty in climate change, land use scenarios, and the timing and patterns of disturbance events, a probabilistic and multimodel approach is the most useful for generating accurate predictions of forest permanence risk with anthropogenic climate change on decadal-to-centennial time scales (*122*). Such an approach has the potential to include the range of uncertainties in future climate, disturbance event characteristics, and human land use and management scenarios. We also posit that the most-credible estimates of forest permanence risks will be those evaluated against observations of disturbances and impacts that include confidence intervals. Continual testing and refinement of mechanistic models against remotely sensed and ground data in an ecological forecasting endeavor has the potential to yield results with higher confidence, similar to improvements in weather forecasting models over the past several decades (*123*).

Remote sensing data

The broad spatial coverage and increasingly long time series of satellite remote sensing data make such datasets highly useful for quantifying permanence risks and informing mechanistic models. The forest research and policy community now has access to a >35-year record of Landsat satellite series observations at 30-m

ILLUSTRATION: DAVID MEIKLE

resolution globally. These datasets have allowed multidecadal assessments of forest losses and gains (*124*) and provide key information on forest disturbance and recovery, yielding insights into the relative permanence of forest landscapes globally. A central remaining challenge is to attribute observed Landsat forest loss and gain to specific types of disturbances.

Although these existing datasets provide a useful framework for monitoring global forests and assessing drivers of change, they are increasingly being augmented by new Earth observations that provide specific information on the structural and functional attributes of forest ecosystems. These include measurements of three-dimensional forest structure using light detection and ranging (lidar) [GEDI; (*125*)], solar-induced fluorescence measurements that provide information on forest photosynthesis, and very high–resolution imaging with near daily temporal revisit from private satellite constellations (*126*). The next decade will bring several new satellite missions that will further enable more-rigorous permanence risk assessment at global scales and will promote robust ecosystem model assessments, benchmarks, and comparisons. For example, two new radar missions, the P-band European Space Agency (ESA) BIOMASS and the L-band NASA-India NISAR will provide, for the first time, coincident space-based multifrequency interferometric measurements of forest structural properties at high (~30 m) spatial resolution globally, including sensitivity in high-biomass regions like the tropics.

Ways forward to bridge the science-policy divide

Tools to leverage the best available science

Rapid advances in global datasets and mechanistic models have the potential to shed light on the future of the land carbon sink and to inform F-NCS policy. New computational methods will likely be needed to integrate data across spatial and temporal scales, blend observational and mechanistic analyses, and forecast uncertainty in statistically rigorous ways. For these advances to then be widely used, they should be wrapped in tools that are openly accessible, transparent, modular, applicable across scales, and usable by a wide range of stakeholders. Global ecologists must continue to expand code and data sharing and the open-access publication of results and to leverage modern cloud technologies for processing and sharing large datasets. At this intersection, there are many opportunities for scientists to partner with the broader software community to improve the performance, documentation, interpretability, and usability of these tools to meet the needs of key stakeholders.

Uptake of new science into policy

For government policy processes to take up the most recent scientific understanding of permanence and other relevant risks, policy-makers need to be aware of and open to new information. The relationship between policy-makers and scientific information is complex (*127*) and may be most challenging when new information raises fundamental questions about policy-makers' prior assumptions (*128*). Frequent and formal review mechanisms within F-NCS policies and the willingness of policy-makers to consider new information—especially information critical of current practices—will ensure the uptake of new scientific findings concerning permanence risks to forest carbon projects. We also urge caution when considering calculations of F-NCS potential that ignore important constraints from biogeochemistry, biophysical feedbacks, timing, and a wide range of human dimensions [e.g., (*129*), which inflated estimates of tree restoration's realistically feasible, climate-cooling carbon storage potential by at least 3- to 10-fold (*130*–*133*)].

In addition to accounting for permanence risks, F-NCS projects must also demonstrate the following: (i) that they are additional (i.e., they reflect climate benefits that would not occur in the absence of the F-NCS project); (ii) that they account for leakage (i.e., climate benefits are calculated to reflect emissions that may be increased elsewhere as a result of the F-NCS project's economic effects on drivers of forest carbon loss); and (iii) that they have a net cooling effect on the climate by accounting for both carbon sequestration—including the full life cycle effects, such as the fate of wood products in the economy—and biophysical impacts of forests on climate through their reflectivity, evapotranspiration, and surface roughness (*15*, *134*). Many of the credits in early carbon offset programs came from projects that were subsequently estimated to be non-additional and therefore likely led to higher net emissions (*135*, *136*). Ongoing evaluation of permanence, additionality, and leakage concerns is critically important for forest offset programs because any overcrediting allows polluters to increase their emissions more than the offset project reduces emissions. In contrast, overcrediting in a public investments or results-based finance framework to protect forest carbon (i.e., not directly offsetting emissions) will only reduce the extent of climate benefits but will not lead to a net societal harm through more greenhouse gases being emitted to the atmosphere.

Outlook

F-NCSs have the potential to contribute to climate mitigation. Crucially, however, F-NCS efforts must not distract from other urgent mitigation activities, particularly major reductions in fossil fuel emissions, and they need to be informed by good science to be successful. Inadequate treatment of permanence carries major risks that disturbance-driven reversals in F-NCS projects could worsen climate change, which is especially dangerous if F-NCSs are used to justify further fossil fuel emissions. The scientific community has a broad array of datasets and tools to estimate and forecast permanence risk of F-NCS projects, which are not widely used in current F-NCS efforts, and these datasets and tools will grow rapidly in coming years. Climate change will fundamentally increase permanence risks to forest ecosystems over the 21st century. An ambitious scientific research agenda that leverages large-scale datasets and mechanistic models has the potential to transform our scientific understanding of the future of Earth's forests and to provide critical policy-relevant information. A broad, multidisciplinary effort that extends beyond scientists is needed to ensure that the best available science is accessible to and used in policy and management approaches.

REFERENCES AND NOTES

1. P. Friedlingstein *et al*., Global carbon budget 2019. *Earth Syst. Sci. Data* **11**, 1783–1838 (2019). doi: 10.5194/essd-11-1783-2019
2. Y. Pan *et al*., A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011). doi: 10.1126/science.1201609; pmid: 21764754
3. T. A. M. Pugh *et al*., Role of forest regrowth in global carbon sink dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4382–4387 (2019). doi: 10.1073/pnas.1810512116; pmid: 30782807
4. K. Thompson, Forests and climate change in America: Some early views. *Clim. Change* **3**, 47–64 (1980). doi: 10.1007/BF02423168
5. R. A. Birdsey, A. J. Plantinga, L. S. Heath, Past and prospective carbon storage in United States forests. *For. Ecol. Manage.* **58**, 33–40 (1993). doi: 10.1016/0378-1127(93)90129-B
6. G. B. Bonan, Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science* **320**, 1444–1449 (2008). doi: 10.1126/science.1155121; pmid: 18556546
7. B. W. Griscom *et al*., Natural climate solutions. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11645–11650 (2017). doi: 10.1073/pnas.1710465114; pmid: 29078344
8. S. Roe *et al*., Contribution of the land sector to a 1.5° C world. *Nat. Clim. Chang.* **9**, 817–828 (2019). doi: 10.1038/s41558-019-0591-9
9. B. W. Griscom *et al*., National mitigation potential from natural climate solutions in the tropics. *Phil. Trans. R. Soc. B* **375**, 20190126 (2020). doi: 10.1098/rstb.2019.0126
10. J. E. Fargione *et al*., Natural climate solutions for the United States. *Sci. Adv.* **4**, eaat1869 (2018). doi: 10.1126/sciadv.aat1869
11. W. R. Turner *et al*., Global conservation of biodiversity and ecosystem services. *Bioscience* **57**, 868–873 (2007). doi: 10.1641/B571009
12. California Air Resources Board, Compliance Offset Program, https://ww3.arb.ca.gov/cc/capandtrade/offsets/offsets.htm [accessed 1 Jan 2020].
13. G. Simonet, A. B. Bos, A. E. Duchelle, I. A. P. Resosudarmo, J. Subervie, S. Wunder, in *Transforming REDD+: Lessons and New Directions*, A. Angelsen, Ed. (Center for International Forestry Research, 2018), pp. 117–130.
14. A. Roopsind, B. Sohngen, J. Brandt, Evidence that a national REDD+ program reduces tree cover loss and carbon emissions in a high forest cover, low deforestation country. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24492–24499 (2019). doi: 10.1073/pnas.1904027116; pmid: 31740591
15. C. M. Anderson, C. B. Field, K. J. Mach, Forest offsets partner climate-change mitigation with conservation. *Front. Ecol. Environ.* **15**, 359–365 (2017). doi: 10.1002/fee.1515
16. S. Solomon, G.-K. Plattner, R. Knutti, P. Friedlingstein, Irreversible climate change due to carbon dioxide emissions. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1704–1709 (2009). doi: 10.1073/pnas.0812721106; pmid: 19179281
17. W. A. Kurz, G. Stinson, G. J. Rampley, C. C. Dymond, E. T. Neilson, Risk of natural disturbances makes future contribution of Canada's forests to the global carbon cycle

highly uncertain. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1551–1555 (2008). doi: 10.1073/pnas.0708133105; pmid: 18230736
18. W. R. L. Anderegg, J. M. Kane, L. D. L. Anderegg, Consequences of widespread tree mortality triggered by drought and temperature stress. *Nat. Clim. Chang.* **3**, 30–36 (2013). doi: 10.1038/nclimate1635
19. P. M. Brando *et al.*, Droughts, Wildfires, and Forest Carbon Cycling: A Pantropical Synthesis. *Annu. Rev. Earth Planet. Sci.* **47**, 555–581 (2019). doi: 10.1146/annurev-earth-082517-010235
20. C. D. Allen *et al.*, A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For. Ecol. Manage.* **259**, 660–684 (2010). doi: 10.1016/j.foreco.2009.09.001
21. H. Hartmann *et al.*, Research frontiers for improving our understanding of drought-induced tree and forest mortality. *New Phytol.* **218**, 15–28 (2018). doi: 10.1111/nph.15048; pmid: 29488280
22. L. E. Aragão, Y. E. Shimabukuro, The incidence of fire in Amazonian forests with implications for REDD. *Science* **328**, 1275–1278 (2010). doi: 10.1126/science.1186925; pmid: 20522775
23. B. D. Amiro *et al.*, Ecosystem carbon dioxide fluxes after disturbance in forests of North America. *J. Geophys. Res.* **115**, G00K02 (2010). doi: 10.1029/2010JG001390
24. J. A. Hicke *et al.*, Effects of biotic disturbances on forest carbon cycling in the United States and Canada. *Glob. Change Biol.* **18**, 7–34 (2012). doi: 10.1111/j.1365-2486.2011.02543.x
25. R. Seidl *et al.*, Forest disturbances under climate change. *Nat. Clim. Chang.* **7**, 395–402 (2017). doi: 10.1038/nclimate3303; pmid: 28861124
26. T. A. M. Pugh, A. Arneth, M. Kautz, B. Poulter, B. Smith, Important role of forest disturbances in the global biomass turnover and carbon sinks. *Nat. Geosci.* **12**, 730–735 (2019). doi: 10.1038/s41561-019-0427-2; pmid: 31478009
27. M. Reichstein *et al.*, Climate extremes and the carbon cycle. *Nature* **500**, 287–295 (2013). doi: 10.1038/nature12350; pmid: 23955228
28. P. Friedlingstein *et al.*, Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *J. Clim.* **27**, 511–526 (2014). doi: 10.1175/JCLI-D-12-00579.1
29. K. Tokimatsu, R. Yasuoka, M. Nishio, Global zero emissions scenarios: The role of biomass energy with carbon capture and storage by forested land use. *Appl. Energy* **185**, 1899–1906 (2017). doi: 10.1016/j.apenergy.2015.11.077
30. M. D. Hurteau, B. A. Hungate, G. W. Koch, Accounting for risk in valuing forest carbon offsets. *Carbon Balance Manag.* **4**, 1 (2009). doi: 10.1186/1750-0680-4-1; pmid: 19149889
31. M. D. Hurteau, B. A. Hungate, G. W. Koch, M. P. North, G. R. Smith, Aligning ecology and markets in the forest carbon cycle. *Front. Ecol. Environ.* **11**, 37–42 (2013). doi: 10.1890/120039
32. D. Purves, S. Pacala, Predictive models of forest dynamics. *Science* **320**, 1452–1453 (2008). doi: 10.1126/science.1155359; pmid: 18556548
33. J. P. Fisk *et al.*, The impacts of tropical cyclones on the net carbon balance of eastern US forests (1851–2000). *Environ. Res. Lett.* **8**, 045017 (2013). doi: 10.1088/1748-9326/8/4/045017
34. W. A. Kurz *et al.*, Mountain pine beetle and forest carbon feedback to climate change. *Nature* **452**, 987–990 (2008). doi: 10.1038/nature06777; pmid: 18432244
35. D. Baldocchi, J. Penuelas, The physics and ecology of mining carbon dioxide from the atmosphere by ecosystems. *Glob. Change Biol.* **25**, 1191–1197 (2019). doi: 10.1111/gcb.14559
36. N. Andela *et al.*, A human-driven decline in global burned area. *Science* **356**, 1356–1362 (2017). doi: 10.1126/science.aal4108; pmid: 28663495
37. E. Chuvieco *et al.*, A new global burned area product for climate assessment of fire impacts. *Glob. Ecol. Biogeogr.* **25**, 619–629 (2016). doi: 10.1111/geb.12440
38. G. R. van der Werf *et al.*, Global fire emissions estimates during 1997-2016. *Earth Syst. Sci. Data* **9**, 697–720 (2017). doi: 10.5194/essd-9-697-2017
39. L. Giglio, J. T. Randerson, G. R. van der Werf, Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *J. Geophys. Res. Biogeosci.* **118**, 317–328 (2013). doi: 10.1002/jgrg.20042
40. J. A. Hicke, A. J. Meddens, C. D. Allen, C. A. Kolden, Carbon stocks of trees killed by bark beetles and wildfire in the western United States. *Environ. Res. Lett.* **8**, 035032 (2013). doi: 10.1088/1748-9326/8/3/035032
41. B. M. Rogers, A. J. Soja, M. L. Goulden, J. T. Randerson, Influence of tree species on continental differences in boreal fires and climate feedbacks. *Nat. Geosci.* **8**, 228–234 (2015). doi: 10.1038/ngeo2352
42. G. R. van der Werf *et al.*, Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmos. Chem. Phys.* **10**, 11707–11735 (2010). doi: 10.5194/acp-10-11707-2010
43. F. Mouillot, C. B. Field, Fire history and the global carbon budget: A 1× 1 fire history reconstruction for the 20th century. *Glob. Change Biol.* **11**, 398–420 (2005). doi: 10.1111/j.1365-2486.2005.00920.x
44. A. P. Williams, J. T. Abatzoglou, Recent advances and remaining uncertainties in resolving past and future climate effects on global fire activity. *Curr. Clim. Change Rep.* **2**, 1–14 (2016). doi: 10.1007/s40641-016-0031-0
45. R. Dutta, A. Das, J. Aryal, Big data integration shows Australian bush-fire frequency is increasing significantly. *R. Soc. open sci.* **3**, 150241 (2016). doi: 10.1098/rsos.150241; pmid: 26998312
46. J. C. White, M. A. Wulder, T. Hermosilla, N. C. Coops, G. W. Hobart, A nationwide annual characterization of 25 years of forest disturbance and recovery for Canada using Landsat time series. *Remote Sens. Environ.* **194**, 303–321 (2017). doi: 10.1016/j.rse.2017.03.035
47. M. Flannigan *et al.*, Global wildland fire season severity in the 21st century. *For. Ecol. Manage.* **294**, 54–61 (2013). doi: 10.1016/j.foreco.2012.10.022
48. M. A. Moritz *et al.*, Climate change and disruptions to global fire activity. *Ecosphere* **3**, 1–22 (2012). doi: 10.1890/ES11-00345.1
49. W. Knorr, A. Arneth, L. Jiang, Demographic controls of future global fire risk. *Nat. Clim. Chang.* **6**, 781–785 (2016). doi: 10.1038/nclimate2999
50. F. Li *et al.*, Historical (1700–2012) global multi-model estimates of the fire emissions from the Fire Modeling Intercomparison Project (FireMIP). *Atmos. Chem. Phys.* **19**, 12545–12567 (2019). doi: 10.5194/acp-19-12545-2019
51. D. M. Bowman, B. P. Murphy, G. J. Williamson, M. A. Cochrane, Pyrogeographic models, feedbacks and the future of global fire regimes. *Glob. Ecol. Biogeogr.* **23**, 821–824 (2014). doi: 10.1111/geb.12180
52. C. D. Allen, D. D. Breshears, N. G. McDowell, On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. *Ecosphere* **6**, 1–55 (2015). doi: 10.1890/ES15-00203.1
53. J. Zscheischler *et al.*, others, A few extreme events dominate global interannual variability in gross primary production. *Environ. Res. Lett.* **9**, 035001 (2014). doi: 10.1088/1748-9326/9/3/035001
54. B. M. Sleeter *et al.*, Effects of 21st-century climate, land use, and disturbances on ecosystem carbon balance in California. *Glob. Change Biol.* **25**, 3334–3353 (2019). doi: 10.1111/gcb.14677; pmid: 31066121
55. A. M. Schwantes *et al.*, Measuring canopy loss and climatic thresholds from an extreme drought along a fivefold precipitation gradient across Texas. *Glob. Change Biol.* **23**, 5120–5135 (2017). doi: 10.1111/gcb.13775; pmid: 28649768
56. N. Brouwers, G. Matusick, K. Ruthrof, T. Lyons, G. Hardy, Landscape-scale assessment of tree crown dieback following extreme drought and heat in a Mediterranean eucalypt forest ecosystem. *Landsc. Ecol.* **28**, 69–80 (2013). doi: 10.1007/s10980-012-9815-3
57. A. Dai, Drought under global warming: A review. *WIREs Clim. Change* **2**, 45–65 (2011). doi: 10.1002/wcc.81
58. W. R. Anderegg *et al.*, Meta-analysis reveals that hydraulic traits explain cross-species patterns of drought-induced tree mortality across the globe. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5024–5029 (2016). doi: 10.1073/pnas.1525678113; pmid: 27091965
59. S. Greenwood *et al.*, Tree mortality across biomes is promoted by drought intensity, lower wood density and higher specific leaf area. *Ecol. Lett.* **20**, 539–553 (2017). doi: 10.1111/ele.12748; pmid: 28220612
60. W. R. L. Anderegg, L. D. L. Anderegg, K. L. Kerr, A. T. Trugman, Widespread drought-induced tree mortality at dry range edges indicates that climate stress exceeds species' compensating mechanisms. *Glob. Change Biol.* **25**, 3793–3802 (2019). doi: 10.1111/gcb.14771; pmid: 31323157
61. B. M. Rogers *et al.*, Detecting early warning signals of tree mortality in boreal North America using multiscale satellite data. *Glob. Change Biol.* **24**, 2284–2304 (2018). doi: 10.1111/gcb.14107; pmid: 29481709
62. M. L. Goulden, R. C. Bales, California forest die-off linked to multi-year deep soil drying in 2012–2015 drought. *Nat. Geosci.* **12**, 632–637 (2019). doi: 10.1038/s41561-019-0388-5
63. A. T. Trugman *et al.*, Tree carbon allocation explains forest drought-kill and recovery patterns. *Ecol. Lett.* **21**, 1552–1560 (2018). doi: 10.1111/ele.13136; pmid: 30125446
64. W. R. L. Anderegg *et al.*, The roles of hydraulic and carbon stress in a widespread climate-induced forest die-off. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 233–237 (2012). doi: 10.1073/pnas.1107891109; pmid: 22167807
65. H. D. Adams *et al.*, A multi-species synthesis of physiological mechanisms in drought-induced tree mortality. *Nat. Ecol. Evol.* **1**, 1285–1291 (2017). doi: 10.1038/s41559-017-0248-x; pmid: 29046541
66. B. Choat *et al.*, Triggers of tree mortality under drought. *Nature* **558**, 531–539 (2018). doi: 10.1038/s41586-018-0240-x; pmid: 29950621
67. M. T. Tyree, J. S. Sperry, Vulnerability of Xylem to Cavitation and Embolism. *Annu. Rev. Plant Phys. Mol. Bio.* **40**, 19–36 (1989). doi: 10.1146/annurev.pp.40.060189.000315
68. A. J. Meddens, J. A. Hicke, C. A. Ferguson, Spatiotemporal patterns of observed bark beetle-caused tree mortality in British Columbia and the western United States. *Ecol. Appl.* **22**, 1876–1891 (2012). doi: 10.1890/11-1785.1; pmid: 23210306
69. C. Senf, M. A. Wulder, E. M. Campbell, P. Hostert, Using Landsat to assess the relationship between spatiotemporal patterns of western spruce budworm outbreaks and regional-scale weather variability. *Can. J. Rem. Sens.* **42**, 706–718 (2016). doi: 10.1080/07038992.2016.1220828
70. M. Kautz, A. J. Meddens, R. J. Hall, A. Arneth, Biotic disturbances in Northern Hemisphere forests–a synthesis of recent data, uncertainties and implications for forest monitoring and modelling. *Glob. Ecol. Biogeogr.* **26**, 533–552 (2017). doi: 10.1111/geb.12558
71. J. U. Jepsen, S. B. Hagen, R. A. Ims, N. G. Yoccoz, Climate change and outbreaks of the geometrids Operophtera brumata and Epirrita autumnata in subarctic birch forest: Evidence of a recent outbreak range expansion. *J. Anim. Ecol.* **77**, 257–264 (2008). doi: 10.1111/j.1365-2656.2007.01339.x; pmid: 18070041
72. D. A. Herms, D. G. McCullough, Emerald ash borer invasion of North America: History, biology, ecology, impacts, and management. *Annu. Rev. Entomol.* **59**, 13–30 (2014). doi: 10.1146/annurev-ento-011613-162051; pmid: 24112110
73. B. J. Bentz *et al.*, Climate Change and Bark Beetles of the Western United States and Canada: Direct and Indirect Effects. *Bioscience* **60**, 602–613 (2010). doi: 10.1525/bio.2010.60.8.6
74. T. E. Kolb *et al.*, Observed and anticipated impacts of drought on forest insects and diseases in the United States. *For. Ecol. Manage.* **380**, 321–334 (2016). doi: 10.1016/j.foreco.2016.04.051
75. C. C. Dymond *et al.*, Future spruce budworm outbreak may create a carbon source in eastern Canadian forests. *Ecosystems* **13**, 917–931 (2010). doi: 10.1007/s10021-010-9364-z
76. R. N. Sturrock *et al.*, Climate change and forest diseases. *Plant Pathol.* **60**, 133–149 (2011). doi: 10.1111/j.1365-3059.2010.02406.x
77. K. F. Raffa *et al.*, Cross-scale Drivers of Natural Disturbances Prone to Anthropogenic Amplification: The Dynamics of Bark Beetle Eruptions. *Bioscience* **58**, 501–517 (2008). doi: 10.1641/B580607
78. S. Trumbore, P. Brando, H. Hartmann, Forest health and global change. *Science* **349**, 814–818 (2015). doi: 10.1126/science.aac6759; pmid: 26293952
79. J. Huang *et al.*, Tree defence and bark beetles in a drying world: Carbon partitioning, functioning and modelling. *New Phytol.* **225**, 26–36 (2020). doi: 10.1111/nph.16173; pmid: 31494935
80. T. L. Shore, L. Safranyik, J. P. Lemieux, Susceptibility of lodgepole pine stands to the mountain pine beetle: Testing of a rating system. *Can. J. For. Res.* **30**, 44–49 (2000). doi: 10.1139/x99-182
81. P. C. Buotte *et al.*, Near-future forest vulnerability to drought and fire varies across the western United States. *Glob. Change Biol.* **25**, 290–303 (2019). doi: 10.1111/gcb.14490; pmid: 30444042
82. A. M. Jönsson, L. M. Schroeder, F. Lagergren, O. Anderbrant, B. Smith, Guess the impact of Ips typographus—An ecosystem modelling approach for simulating spruce bark beetle outbreaks. *Agric. For. Meteorol.* **166–167**, 188–200 (2012). doi: 10.1016/j.agrformet.2012.07.012

83. J. Q. Chambers *et al*., Hurricane Katrina's carbon footprint on U.S. Gulf Coast forests. *Science* **318**, 1107 (2007). doi: 10.1126/science.1148913; pmid: 18006740
84. L. R. Iverson *et al*., Multi-model comparison on the effects of climate change on tree species in the eastern US: Results from an enhanced niche model and process-based ecosystem and landscape models. *Landsc. Ecol.* **32**, 1327–1346 (2017). doi: 10.1007/s10980-016-0404-8
85. T. Zhang, Ü. Niinemets, J. Sheffield, J. W. Lichstein, Shifts in tree functional composition amplify the response of forest biomass to climate. *Nature* **556**, 99–102 (2018). doi: 10.1038/nature26152; pmid: 29562235
86. A. T. Trugman, L. D. L. Anderegg, J. D. Shaw, W. R. L. Anderegg, Trait velocities reveal that mortality has driven widespread coordinated shifts in forest hydraulic trait composition. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8532–8538 (2020). doi: 10.1073/pnas.1917521117; pmid: 32229563
87. M. J. Duveneck, J. R. Thompson, E. J. Gustafson, Y. Liang, A. M. de Bruijn, Recovery dynamics and climate change effects to future New England forests. *Landsc. Ecol.* **32**, 1385–1397 (2017). doi: 10.1007/s10980-016-0415-5
88. G. Tang, B. Beckage, B. Smith, Potential future dynamics of carbon fluxes and pools in New England forests and their climatic sensitivities: A model-based study. *Global Biogeochem. Cycles* **28**, 286–299 (2014). doi: 10.1002/2013GB004656
89. L. A. Brandt *et al*., Integrating science and management to assess forest ecosystem vulnerability to climate change. *J. For.* **115**, 212–221 (2017). doi: 10.5849/jof.15-147
90. C. Swanston *et al*., Vulnerability of forests of the Midwest and Northeast United States to climate change. *Clim. Change* **146**, 103–116 (2018). doi: 10.1007/s10584-017-2065-2
91. K. M. Miller, B. J. McGill, Compounding human stressors cause major regeneration debt in over half of eastern US forests. *J. Appl. Ecol.* **56**, 1355–1366 (2019). doi: 10.1111/1365-2664.13375
92. K. T. Davis *et al*., Wildfires and climate change push low-elevation forests across a critical climate threshold for tree regeneration. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 6193–6198 (2019). doi: 10.1073/pnas.1815107116; pmid: 30858310
93. C. D. Allen, Interactions across spatial scales among forest dieback, fire, and erosion in northern New Mexico landscapes. *Ecosystems* **10**, 797–808 (2007). doi: 10.1007/s10021-007-9057-4
94. J. M. Kane, J. M. Varner, M. R. Metz, P. J. van Mantgem, Characterizing interactions between fire and other disturbances and their impacts on tree mortality in western US Forests. *For. Ecol. Manage.* **405**, 188–199 (2017). doi: 10.1016/j.foreco.2017.09.037
95. W. R. Anderegg *et al*., Tree mortality from drought, insects, and their interactions in a changing climate. *New Phytol.* **208**, 674–683 (2015). doi: 10.1111/nph.13477; pmid: 26058406
96. N. L. Stephenson, A. J. Das, N. J. Ampersee, B. M. Bulaon, J. L. Yee, Which trees die during drought? The key role of insect host-tree selection. *J. Ecol.* **107**, 2383–2401 (2019). doi: 10.1111/1365-2745.13176
97. F. Krausmann, K.-H. Erb, S. Gingrich, C. Lauk, H. Haberl, Global patterns of socioeconomic biomass flows in the year 2000: A comprehensive assessment of supply, consumption and constraints. *Ecol. Econ.* **65**, 471–487 (2008). doi: 10.1016/j.ecolecon.2007.07.012
98. M. Gren, A. Z. Aklilu, Policy design for forest carbon sequestration: A review of the literature. *For. Policy Econ.* **70**, 128–136 (2016). doi: 10.1016/j.forpol.2016.06.008
99. S. M. Hood, S. Baker, A. Sala, Fortifying the forest: Thinning and burning increase resistance to a bark beetle outbreak and promote forest resilience. *Ecol. Appl.* **26**, 1984–2000 (2016). doi: 10.1002/eap.1363; pmid: 27755724
100. D. A. Peltzer, R. B. Allen, G. M. Lovett, D. Whitehead, D. A. Wardle, Effects of biological invasions on forest carbon sequestration. *Glob. Change Biol.* **16**, 732–746 (2010). doi: 10.1111/j.1365-2486.2009.02038.x
101. J. T. Abatzoglou, A. P. Williams, Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11770–11775 (2016). doi: 10.1073/pnas.1607171113; pmid: 27791053
102. E. Marland *et al*., *Understanding and Analysis: The California Air Resources Board Forest Offset Protocol* (Springer, 2017).
103. California Air Resources Board, Compliance Offset Protocol U.S. Forest Offset Projects (2014); https://ww3.arb.ca.gov/cc/capandtrade/protocols/usforest/usforestprojects_2014.htm.
104. California Air Resources Board, Q4 2019 Compliance Instrument Report (2019); https://ww2.arb.ca.gov/compliance-instrument-report.
105. Certification Center on Climate Change, Japan, Forest Carbon Sink Becomes Carbon Offsetting Credit (2009); www.env.go.jp/en/earth/ets/mkt_mech/fcsb-coc.pdf.
106. D. Diaz, "Moving Beyond the Buffer Pool" (Ecosystem Marketplace, 2010); www.ecosystemmarketplace.com/articles/moving-beyond-the-buffer-pool/.
107. Y. Malhi *et al*., An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR). *J. Veg. Sci.* **13**, 439–450 (2002). doi: 10.1111/j.1654-1103.2002.tb02068.x
108. K. J. Anderson-Teixeira *et al*., CTFS-ForestGEO: A worldwide network monitoring forests in an era of global change. *Glob. Change Biol.* **21**, 528–549 (2015). doi: 10.1111/gcb.12712; pmid: 25258024
109. B. M. Sanderson, R. A. Fisher, A fiery wake-up call for climate science. *Nat. Clim. Chang.* **10**, 175–177 (2020). doi: 10.1038/s41558-020-0707-2
110. R. A. Fisher *et al*., Vegetation demographics in Earth System Models: A review of progress and priorities. *Glob. Change Biol.* **24**, 35–54 (2018). doi: 10.1111/gcb.13910; pmid: 28921829
111. H. Bugmann, A review of forest gap models. *Clim. Change* **51**, 259–305 (2001). doi: 10.1023/A:1012525626267
112. D. Medvigy, S. C. Wofsy, J. W. Munger, D. Y. Hollinger, P. R. Moorcroft, Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. *J. Geophys. Res.* **114**, G01002 (2009). doi: 10.1029/2008JG000812
113. H. Bugmann *et al*., Tree mortality submodels drive simulated long-term forest dynamics: Assessing 15 models from the stand to global scale. *Ecosphere* **10**, e02616 (2019). doi: 10.1002/ecs2.2616
114. K. Yu *et al*., Pervasive decreases in living vegetation carbon turnover time across forest climate zones. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24662–24667 (2019). doi: 10.1073/pnas.1821387116; pmid: 31740604
115. N. G. McDowell *et al*., The interdependence of mechanisms underlying climate-driven vegetation mortality. *Trends Ecol. Evol.* **26**, 523–532 (2011). doi: 10.1016/j.tree.2011.06.003; pmid: 21802765
116. X. Xu, D. Medvigy, J. S. Powers, J. M. Becknell, K. Guan, Diversity in plant hydraulic traits explains seasonal and inter-annual variations of vegetation dynamics in seasonally dry tropical forests. *New Phytol.* **212**, 80–95 (2016). doi: 10.1111/nph.14009; pmid: 27189787
117. D. Kennedy *et al*., Implementing plant hydraulics in the Community Land Model, version 5. *J. Adv. Model. Earth Syst.* **11**, 485–513 (2019). doi: 10.1029/2018MS001500
118. J. S. Sperry *et al*., The impact of rising CO_2 and acclimation on the response of US forests to global warming. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25734–25744 (2019). doi: 10.1073/pnas.1913072116; pmid: 31767760
119. S. Kloster *et al*., Fire dynamics during the 20th century simulated by the Community Land Model. *Biogeosci.* **7**, 1877–1902 (2010). doi: 10.5194/bgd-7-565-2010
120. M. C. Dietze, J. H. Matthes, A general ecophysiological framework for modelling the impact of pests and pathogens on forest ecosystems. *Ecol. Lett.* **17**, 1418–1426 (2014). doi: 10.1111/ele.12345; pmid: 25168168
121. Y. Luo, T. F. Keenan, M. Smith, Predictability of the terrestrial carbon cycle. *Glob. Change Biol.* **21**, 1737–1751 (2015). doi: 10.1111/gcb.12766; pmid: 25327167
122. D. N. Huntzinger *et al*., Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions. *Sci. Rep.* **7**, 4765 (2017). doi: 10.1038/s41598-017-03818-2; pmid: 28684755
123. M. C. Dietze *et al*., Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1424–1432 (2018). doi: 10.1073/pnas.1710231115; pmid: 29382745
124. M. C. Hansen *et al*., High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850–853 (2013). doi: 10.1126/science.1244693
125. R. Dubayah *et al*., The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Sci. Remote Sens.* **1**, 100002 (2020). doi: 10.1016/j.srs.2020.100002
126. D. Schimel, F. D. Schneider, JPL Carbon and Ecosystem Participants, Flux towers in the sky: Global ecology from space. *New Phytol.* **224**, 570–584 (2019). doi: 10.1111/nph.15934; pmid: 31112309
127. J. Miller, Building a better dialogue between energy research and policy. *Nat. Energy* **4**, 816–818 (2019). doi: 10.1038/s41560-019-0483-2
128. B. Haya, "Policy brief: The California Air Resources Board's U.S. Forest offset protocol underestimates leakage" (2019); https://gspp.berkeley.edu/assets/uploads/research/pdf/Policy_Brief-US_Forest_Projects-Leakage-Haya_2.pdf.
129. J.-F. Bastin *et al*., The global tree restoration potential. *Science* **365**, 76–79 (2019). doi: 10.1126/science.aax0848; pmid: 31273120
130. J. W. Veldman *et al*., Comment on "The global tree restoration potential". *Science* **366**, eaay7976 (2019). doi: 10.1126/science.aay7976; pmid: 31624182
131. A. K. Skidmore, T. Wang, K. de Bie, P. Pilesjö, Comment on "The global tree restoration potential". *Science* **366**, eaaz0111 (2019). doi: 10.1126/science.aaz0111; pmid: 31780528
132. S. L. Lewis, E. T. A. Mitchard, C. Prentice, M. Maslin, B. Poulter, Comment on "The global tree restoration potential". *Science* **366**, eaaz0388 (2019). doi: 10.1126/science.aaz0388; pmid: 31624179
133. P. Friedlingstein, M. Allen, J. G. Canadell, G. P. Peters, S. I. Seneviratne, Comment on "The global tree restoration potential". *Science* **366**, eaay8060 (2019). doi: 10.1126/science.aay8060; pmid: 31624183
134. R. G. Anderson *et al*., Biophysical considerations in forestry for climate protection. *Front. Ecol. Environ.* **9**, 174–182 (2011). doi: 10.1890/090179
135. L. Schneider, A. Kollmuss, Perverse effects of carbon markets on HFC-23 and SF 6 abatement projects in Russia. *Nat. Clim. Chang.* **5**, 1061–1063 (2015). doi: 10.1038/nclimate2772
136. M. Cames, R. O. Harthan, J. Füssler, M. Lazarus, C. M. Lee, P. Erickson, R. Spalding-Fecher, How additional is the clean development mechanism (DG CLIMA report, 2016); https://ec.europa.eu/clima/sites/clima/files/ets/docs/clean_dev_mechanism_en.pdf.
137. K. Hamrick, M. Gallant, Fertile Ground: State of Forest Carbon Finance 2017 (2017); www.forest-trends.org/publications/fertile-ground/.

ACKNOWLEDGMENTS

We thank the reviewers for insightful comments that improved the paper. **Funding:** W.R.L.A. acknowledges funding from the David and Lucille Packard Foundation; NSF grants 1714972 and 1802880; and the U.S. Department of Agriculture (USDA) National Institute of Food and Agriculture, Agricultural and Food Research Initiative Competitive Programme, Ecosystem Services and Agro-ecosystem Management, grant no. 2018-67019-27850. A.T.T. acknowledges funding from the USDA National Institute of Food and Agriculture, Agricultural and Food Research Initiative Competitive Programme grant no. 2018-67012-31496 and the University of California Laboratory Fees Research Program award no. LFR-20-652467. S.J.G. acknowledges support from NASA Earth Ventures grant NNL15AA03C and NASA Applied Sciences grant NNX17AG51G. J.A.H. acknowledges support by the NSF under grant no. DMS-1520873. J.T.R. acknowledges support from the U.S. Department of Energy Office of Science Biological and Environmental Research RUBISCO science focus area and the University of California - National Laboratory laboratory fees program. D.C. is a member of California's Independent Emissions Market Advisory Committee but does not speak for the Committee here. P.C. acknowledges support from the European Research Council Synergy project SyG-2013-610028 IMBALANCE-P and the ANR CLAND Convergence Institute. R.B.J. acknowledges support from the Andrew W. Mellon Foundation (GBMF5439). **Author contributions:** W.R.L.A., A.T.T., and G.B. designed the project. W.R.L.A., G.B., and J.F. provided data visualizations. All authors participated in the generating of ideas, writing of initial drafts, and revision of subsequent drafts. **Competing interests:** The authors declare no competing interests.

10.1126/science.aaz7005

RESEARCH ARTICLE SUMMARY

SOLAR CELLS

Gas chromatography–mass spectrometry analyses of encapsulated stable perovskite solar cells

Lei Shi*, Martin P. Bucknall, Trevor L. Young, Meng Zhang, Long Hu, Jueming Bing, Da Seul Lee, Jincheol Kim, Tom Wu, Noboru Takamure, David R. McKenzie, Shujuan Huang, Martin A. Green, Anita W. Y. Ho-Baillie*

INTRODUCTION: Although advances in materials and processing have led to remarkable advancements in the energy conversion efficiency of perovskite solar cells (PSCs), increasing from 3.8% to 25.2% in only 10 years, these solar cells cannot become commercially viable unless their underperforming durability is improved. The instability of perovskites must be addressed if PSCs are to compete with silicon technology, which currently offers a 25-year performance warranty. Previous approaches to this problem include the use of metal oxide barrier layers and butyl rubber sealants. Here, we report a low-cost polymer/glass stack encapsulation scheme that enables PSCs to pass the demanding International Electrotechnical Commission (IEC) 61215:2016 Damp Heat and Humidity Freeze tests. These tests help to determine whether solar cell modules can withstand the effects of outdoor operating conditions by exposing them to repeated temperature cycling (–40° to 85°C) as well as 85% relative humidity. Our airtight encapsulation scheme prevented moisture ingress. It was also effective in suppressing outgassing of decomposition products, which limits decomposition reactions of organic hybrid PSCs by allowing these reactions to come to equilibrium. The gas compositions were verified by gas chromatography–mass spectrometry (GC-MS).

RATIONALE: In the GC-MS technique, gas chromatography separates the components in a mixture, and the chemical identity of each component is determined with mass spectrometry. We could directly identify with high specificity the decomposition products of multi-cation perovskite precursors, of unencapsulated perovskite test structures, and of encapsulated full cells at elevated temperatures. The results allowed us to identify thermal degradation pathways by determining the outgassing products of mixed-cation perovskites during heating. We then used GC-MS to evaluate the effectiveness of different packaging techniques developed for PSCs. The packaging schemes were a polyisobutylene (PIB)–based polymer blanket encapsulation, a polyolefin-based blanket encapsulation, and a PIB edge seal. These packaging layers were then capped by a glass cover. For the edge seal, the decomposition gases inside the cell were sampled with a syringe. The feasibilities of these packaging techniques were also demonstrated by IEC photovoltaic module standard Damp Heat and Humidity Freeze testing.

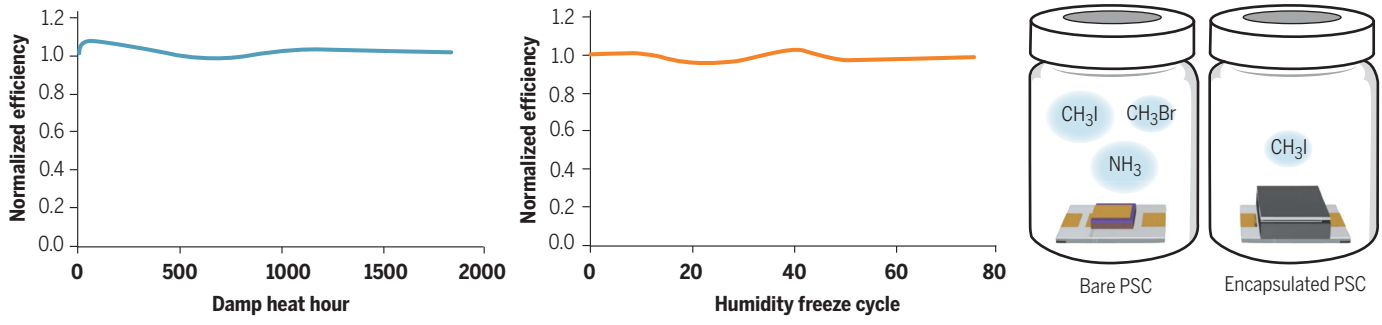
RESULTS: Signature decomposition products such as CH_3I, CH_3Br, and NH_3 were identified and decomposition pathways were proposed for CH_3NH_3I (MAI), $HC(NH_2)_2I$ (FAI), CH_3NH_3Br (MABr), and mixed-cation and mixed-halide $(FAI)_{0.85} + (MABr)_{0.15}$ perovskite precursors, including their secondary decomposition reactions at 350°, 140°, and 85°C. The GC-MS results confirmed that the Br-containing precursor was less prone to thermal decomposition than an I-containing precursor. Also, CsFAMA cells were found to outgas one-fifth as much decomposition product as their FAMA counterparts, which indicated that the Cs-containing cells had better thermal stability. Although the decomposition of FAI is reversible, the mixing of MA with FA precursors caused decomposition products to participate in the secondary reaction that was irreversible. This finding confirmed the disadvantage of mixing of MA with FA perovskite through the reduction in chemical stability. The blanket-encapsulated PSCs sustained no efficiency degradation after 1800 hours of Damp Heat testing or 75 cycles of Humidity Freeze testing.

CONCLUSION: GC-MS identified signature volatile products of the decomposition of organic hybrid perovskites under thermal stress, thereby informing decomposition pathways. The findings are important for developing potential cell-stabilizing strategies, given that cells in the field typically experience high operating temperatures. In addition, results of GC-MS confirm that the low-cost pressure-tight encapsulation we developed is effective in suppressing such outgassing and therefore decomposition reactions of PSCs. This encapsulation scheme is the simplest of all for perovskite cells to pass IEC photovoltaic module standard tests. Our approach can be applied to evaluating the effectiveness of other packaging approaches, as well as testing the effectiveness of coatings and material compositions aimed at limiting light and thermal degradation. ■

ON OUR WEBSITE

Read the full article at https://dx.doi.org/10.1126/science.aba2412

The list of author affiliations is available in the full article online.
*Corresponding author. Email: anita.ho-baillie@sydney.edu.au (A.W.Y.H.-B.); lei.shi@unsw.edu.au (L.S.)
Cite this article as L. Shi *et al.*, *Science* **368**, eaba2412 (2020). DOI: 10.1126/science.aba2412



Stable perovskite solar cells exceeding the requirements of the IEC 61215 Damp Heat and Humidity Freeze tests. Unencapsulated and encapsulated perovskite cells were analyzed by gas chromatography–mass spectrometry, detecting signature volatile products of the organic hybrid perovskite decomposition under thermal stress and confirming the effectiveness of the low-cost pressure-tight polymer/glass stack encapsulation schemes developed.

RESEARCH ARTICLE

SOLAR CELLS

Gas chromatography–mass spectrometry analyses of encapsulated stable perovskite solar cells

Lei Shi[1]*, Martin P. Bucknall[2,3], Trevor L. Young[1], Meng Zhang[1,4], Long Hu[5], Jueming Bing[1,6,7], Da Seul Lee[1], Jincheol Kim[1,8], Tom Wu[5], Noboru Takamure[6], David R. McKenzie[6], Shujuan Huang[1,9], Martin A. Green[1], Anita W. Y. Ho-Baillie[1,6,7]*

Although perovskite solar cells have produced remarkable energy conversion efficiencies, they cannot become commercially viable without improvements in durability. We used gas chromatography–mass spectrometry (GC-MS) to reveal signature volatile products of the decomposition of organic hybrid perovskites under thermal stress. In addition, we were able to use GC-MS to confirm that a low-cost polymer/glass stack encapsulation is effective in suppressing such outgassing. Using such an encapsulation scheme, we produced multi-cation, multi-halide perovskite solar cells containing methylammonium that exceed the requirements of the International Electrotechnical Commission 61215:2016 standard by surviving more than 1800 hours of the Damp Heat test and 75 cycles of the Humidity Freeze test.

Metal halide perovskite solar cells (PSCs) have undergone substantial improvements in power conversion efficiency (PCE), which has increased from 3.8% to 25.2% in only 10 years (*1*–*4*). Coupled with its potential for low fabrication cost (*5*), PSC technology is very promising either as a stand-alone perovskite cell or as a top cell in a tandem device. For PSCs to be commercially viable, however, they need to endure long-term environmental stresses imposed by moisture (*6*–*12*), heat (*12*–*17*), and light (*17*–*19*). Over the years, industry and research groups have developed technologies for moisture barriers (*20*–*24*) such as epoxy, butyl rubber, ceramic thin films, and dyads, which can effectively isolate perovskite and organic solar cells from high-humidity environments (*25*–*29*). Other strategies of varying cell design and material choice to improve moisture and thermal stability include altering three-dimensional (3D) perovskite composition (*30*–*34*), applying 2D perovskite material (*35*–*39*), and using alternative electron/hole transport materials (*7*, *40*–*43*). Light stability is also effectively improved by some of these methods (*34*, *42*, *44*–*47*).

One strategy to improve the thermal stability of PSCs is to suppress stress-induced gaseous decomposition of the perovskite materials, which can be achieved by a pressure-tight coverage or encapsulation using materials such as Al_2O_3, aluminum zinc oxide, and SnO_2 (*48*–*52*). Lee *et al.* deposited four layers of organic/inorganic dyads (thickness >1 μm) by initiated chemical vapor deposition and atomic layer deposition (ALD) onto $(FAPbI_3)_{0.87}(MAPbBr_3)_{0.13}$ PSCs [FA = formamidinium; MA = methylammonium (CH_3NH_3)], which maintained 95% of their initial efficiency after 300 hours of storage in the dark at 50°C and 50% relative humidity (RH) (*50*). Bush *et al.* used an ALD SnO_2/zinc tin oxide bilayer and a sputtered indium tin oxide (ITO) layer as the barrier and contact layer, respectively; a $Cs_{0.17}FA_{0.83}Pb(I_{0.83}Br_{0.17})_3$ cell, when encapsulated by glass/ethylene vinyl acetate (EVA)/glass and edge-sealed by butyl rubber, endured 1000 hours at 85°C and 85% RH (*46*). This was the first PSC reported to pass the International Electrotechnical Commission (IEC) 61215:2016 Damp Heat test. In their work, cesium-formamidinium (Cs-FA) was used as the mixed cation because of its thermal stability (*32*). However, a mixed-cation PSC able to pass this test would be more desirable if it contained MA because it offers higher flexibility for composition engineering and because MA-containing mixed-cation PSCs have higher performance potential (*53*). Boyd *et al.* managed to stabilize $MAPbI_3$ PSCs by using a similar SnO_2/ITO barrier layer where ITO not only covers the PSC but "wraps around" the cell, such that the barrier stack covers the perovskite edges to inhibit the escape of volatile species from decomposed $MAPbI_3$ perovskite. As a result, cells were stable at 85°C for 1000 hours in inert atmosphere without any additional encapsulation (*49*). These barrier layers, however, require vacuum processes and their deposition conditions must be tuned such that they do not affect the performance of the PSC; such factors inevitably increase the fabrication cost.

In terms of characterizing heat-induced mass loss from $MAPbI_3$, thermal gravimetric analysis (TGA) has shown that $MAPbI_3$ reversibly decomposes to methylamine (CH_3NH_2) and hydrogen iodide (HI), leaving PbI_2 as the solid remnant (*54*, *55*). Coupling thermal analysis with Fourier transform infrared spectroscopy (FTIR), Williams *et al.* showed that MAI (CH_3NH_3I) does not release gaseous products until 240°C (*56*). No methylamine (CH_3NH_2) was detected, but CH_3I and NH_3 were found (*56*, *57*). Given that their samples were perovskite precursor solutions, the findings were convoluted with the dimethylformamide (DMF) solvent signals. Juarez-Perez *et al.* used TGA–mass spectrometry (MS) to investigate the mass loss of solvent-free $MAPbI_3$ single crystals and determined that the gaseous products released are a mixture of CH_3NH_2, HI, CH_3I, and NH_3 (*58*). They also (*59*) found that $MAPbI_3$ and $MAPbBr_3$ thin films started thermal decomposition within an hour of annealing in high vacuum (~10^{-6} torr) at typical solar cell operation temperatures (40° to 60°C). However, the possible presence of DMF solvent residue might have interfered with the results, as its mass spectrum overlaps with those of some of the decomposition products. It is challenging for FTIR and MS to unambiguously identify unknown compounds or multiple known compounds in a mixture (e.g., perovskites with encapsulants) that are detected simultaneously. By contrast, for TGA, the mass loss that led to the cell efficiency drop could be well below the detection limit because perovskite decomposition can start at surfaces or interfaces and proceeds in a layer-by-layer manner (*60*). More recently, Juarez-Perez *et al.* used microwave rotational spectroscopy to unambiguously detect CH_3I and CH_3NH_2 from $MAPbI_3$ decomposition at 280°C (*61*), identifying more than one decomposition pathway. However, HI and NH_3 could not be detected by their spectroscopy setup because the lowest transitions for these two species were beyond the upper frequency limit.

Although much of the reported PSC stability research was conducted at 85°C (table S1), photovoltaic modules in real-world outdoor conditions must be able to withstand a combination of moisture ingress and temperature extremes. The IEC 61215:2016 Damp Heat, Thermal Cycling, and Humidity Freeze tests (*62*) (Table 1) are among the standard accelerated tests specified for commercial modules. Such modules must survive these tests with less than

[1]Australian Centre for Advanced Photovoltaics, School of Photovoltaic and Renewable Energy Engineering, University of New South Wales, Sydney, NSW 2052, Australia. [2]Bioanalytical Mass Spectrometry Facility, Mark Wainwright Analytical Centre, University of New South Wales, Sydney, NSW 2052, Australia. [3]School of Optometry and Vision Science, University of New South Wales, Sydney, NSW 2052, Australia. [4]Institute of Photovoltaics, Southwest Petroleum University, Chengdu 610500, China. [5]School of Materials Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia. [6]School of Physics, University of Sydney, Sydney, NSW 2006, Australia. [7]University of Sydney Nano Institute, University of Sydney, Sydney, NSW 2006, Australia. [8]New and Renewable Energy Research Center, Korea Electronics Technology Institute, Seong-Nam, Republic of Korea. [9]School of Engineering, Macquarie University, Sydney, NSW 2109, Australia.
*Corresponding author. Email: anita.ho-baillie@sydney.edu.au (A.W.Y.H.-B.); lei.shi@unsw.edu.au (L.S.)

Table 1. Specifications of the three IEC 61215:2016 tests.

Test	Conditions	Min. T_{95}
Damp Heat	85°C, 85% RH	1000 hours
Thermal Cycling	−40°C (15 min dwell) to 85°C (15 min dwell) Ramp rate of 100°C/hour	200 cycles
Humidity Freeze	50 rounds of Thermal Cycling as prerequisite, followed by −40°C (30 min dwell) to 85°C, 85% RH (20 hours dwell) Ramp rate of 100°C/hour for 0°C ↔ 85°C Ramp rate of 200°C/hour for 0°C ↔ −40°C	10 cycles

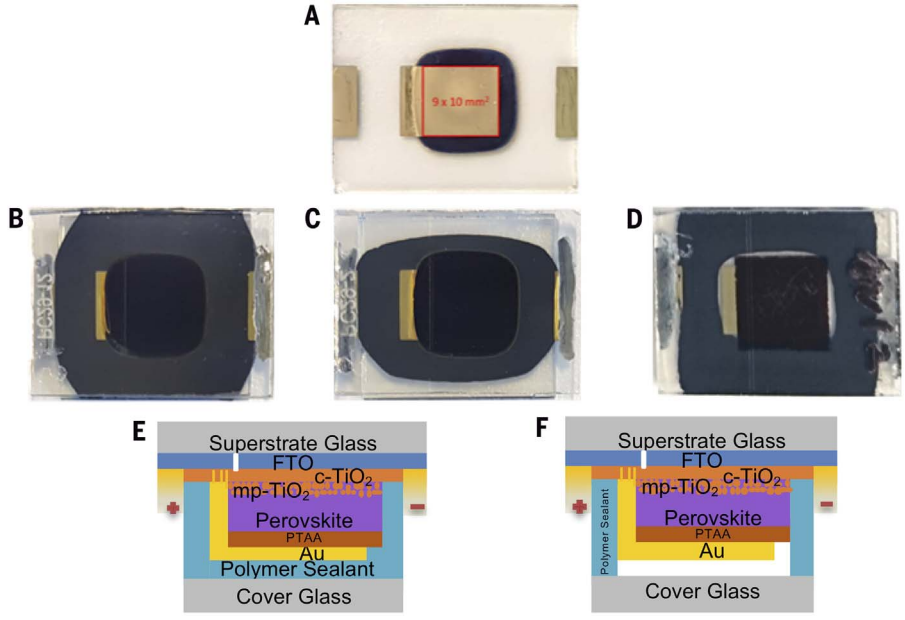


Fig. 1. Solar cells for IEC 61215:2016 Damp Heat and Thermal Cycling tests. (**A**) "Rear" or metal-side view of PSC before packaging (red square denotes the active area). (**B** to **D**) "Front" view (from the superstrate side) of PSC after PIB-based blanket encapsulation (B), PO-based blanket encapsulation (C), and PIB edge seal (D). (**E** and **F**) Illustrations of the cross sections of the respective encapsulation schemes (not to scale).

5% relative loss in efficiency. The Humidity Freeze test, a combination of Damp Heat and Thermal Cycling, therefore exerts the harshest stress on solar cells and their encapsulation. For example, ingressed moisture inside the encapsulation will solidify during freezing and subsequently delaminate the encapsulant. Temperature cycling can promote the decomposition of perovskites (*58*), which itself can accelerate the decomposition reaction. Therefore, weakness in the encapsulation or instability in the perovskite material, if they exist, will most likely be revealed by the Humidity Freeze test.

Previously, we reported a low-cost and effective polyisobutylene (PIB)–based polymeric "blanket-cover" encapsulation scheme for planar $FAPbI_3$ cells passing the IEC 61215:2016 Thermal Cycling test (200 cycles of oscillation between −40° and 85°C) that survived 500 hours in Damp Heat without degradation (*25*). We found that the edge-seal packaging scheme is less effective than the blanket-cover approach and hypothesized that the success of the latter is due to its ability to suppress the outgassing of volatile decomposition products. Here, we developed an improved packaging sequence and applied it to $Cs_{0.05}FA_{0.8}MA_{0.15}Pb(I_{0.85}Br_{0.15})_3$ and $FA_{0.85}MA_{0.15}Pb(I_{0.85}Br_{0.15})_3$. The resulting MA-containing perovskite cells were able to pass the IEC6 1215:2016 Damp Heat and Humidity Freeze tests.

Additionally, we used gas chromatography–mass spectrometry (GC-MS) to directly identify with high specificity the decomposition products of perovskite precursors, unencapsulated and encapsulated perovskite test structures, and full cells at elevated temperature. The findings confirm the outgassing behavior of multication mixed halide perovskites and identify thermal degradation pathways. We then used GC-MS to quantitatively confirm the effectiveness of the polymer blanket cover in conjunction with glass in suppressing such outgassing. This encapsulation scheme demonstrates the feasibility of developing a simple, low-cost packaging technique for stable perovskite cells.

PIB encapsulated and edge-sealed cells

The solar cell structure used was FTO/c-TiO_2/mp-TiO_2/perovskite/PTAA/Au (FTO = fluorine-doped tin oxide; c-TiO_2 = compact TiO_2; mp-TiO_2 = mesoporous TiO_2; PTAA = poly[bis(4-phenyl)(2,4,6-trimethylphenyl)amine]) (Fig. 1, E and F). The composition of the perovskite layer was either $Cs_{0.05}FA_{0.8}MA_{0.15}Pb(I_{0.85}Br_{0.15})_3$ (i.e., CsFAMA) or $FA_{0.85}MA_{0.15}Pb(I_{0.85}Br_{0.15})_3$ (i.e., FAMA).

Figure 1A shows the view from the "rear" or Au side of cell before packaging; Fig. 1, B to D, shows the "front" of the cell (viewing from the superstrate side) after PIB-based polymer (PVS 101) blanket encapsulation, polyolefin (PO)–based polymer blanket encapsulation, and PIB edge seal, respectively. Figure 1, E and F, illustrates the cross sections of the respective encapsulation schemes. Note that the PO used in this work is slightly narrower than the PIB used. The effect of the width of the polymer tape on cell durability is discussed below.

PCE was maintained after encapsulation (table S4). The packaged devices were then subjected to the IEC Damp Heat and Humidity Freeze tests (as specified in Table 1) in an environmental chamber (fig. S2B). The cells' PCEs were measured ex situ at regular intervals. Results are plotted in Fig. 2 and summarized in Table 2.

Regarding cells that underwent the Damp Heat test, PIB-based blanket-encapsulated cells sustained no degradation after 1800 hours, far exceeding the requirement specified by the IEC 61215:2016 standard (1000 hours). The short-circuit current density (J_{SC}) and open circuit voltage (V_{OC}) remained reasonably stable throughout the test, whereas the fill factor (FF) increased by more than 10% relative (fig. S3, A to C). Negligible change in the x-ray diffraction pattern of the CsFAMA perovskite film was observed after 1800 hours of Damp Heat (fig. S3D). On the other hand, PCEs of PIB edge-sealed cells degraded by 30% in the first 200 hours, which is consistent with the observations of edge-sealed $FAPbI_3$ devices in our previous work (*25*). We believe that such a degradation is due to heat-induced decomposition rather than moisture ingress (see below).

Regarding cells that underwent the Humidity Freeze test, PIB-based blanket-encapsulated cells also passed the IEC 61215:2016 test; in particular, the CsFAMA cells experienced no

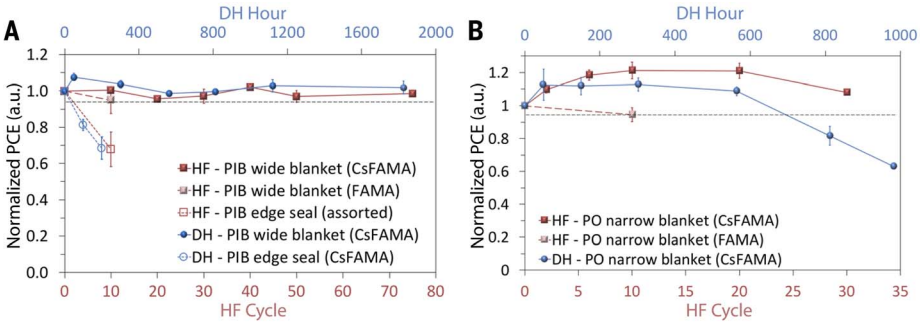


Fig. 2. Results of the IEC 61215:2016 tests demonstrating the effectiveness of packaging schemes. (**A** and **B**) PCE evolution of PIB-encapsulated or edge-sealed PSCs (A) and PO-encapsulated PSCs (B) during Damp Heat (DH) and Humidity Freeze (HF) tests. The dashed lines denote 5% relative PCE loss, which is the maximum allowed by the IEC standard after 1000 hours of Damp Heat or 10 cycles of Humidity Freeze. Error bars denote SD. See Table 2 for absorber composition, quantity, and result summary.

Table 2. Summary of results of Damp Heat and Thermal Cycling tests for PSCs edge-sealed or encapsulated using three different methods.

Test	Encapsulation	No. of samples	Perovskite	T_{95}	Result
Damp Heat	PIB wide blanket	3	CsFAMA	>1800 hours	Pass
	PO narrow blanket	3	CsFAMA	>564 hours	Fail
	PIB edge seal	4	CsFAMA	<100 hours	Fail
Humidity Freeze	PIB wide blanket	4	CsFAMA	>75 cycles	Pass
	PIB wide blanket	2	FAMA	10 cycles	Pass
	PO narrow blanket	4	CsFAMA	>30 cycles	Pass
	PO narrow blanket	3	FAMA	10 cycles	Pass
	PIB edge seal	4	CsFAMA × 2, FAMA × 2	<<10 cycles	Fail

degradation after 75 Humidity Freeze cycles, which is 7.5 times the number required by the IEC 61215:2016 standard. Their J_{SC}, V_{OC}, and FF values were also stable throughout the test (fig. S3, A to C). Interestingly, FAMA cells also passed the 10 cycles required for such tests (taking into account the 50 prerequisite rounds of Thermal Cycling), with average PCE dropping by only 5% after the test. Similar to the Damp Heat test result, PIB edge-sealed PSCs failed the Humidity Freeze test, suffering 30% PCE degradation after 10 Humidity Freeze cycles.

PO encapsulated cells

As edge sealing is the less ideal packaging method, only the blanket approach was tested for PO-encapsulated PSCs. However, the width of the PO tape available from the supplier is narrower than that of the PIB tape. Although the PCEs of PO-encapsulated PSCs increased initially during the Damp Heat test, they decreased after 600 hours. Both J_{SC} and FF contributed to PCE loss (fig. S4, A and C). The decomposition of the perovskite layer induced by the gradual moisture ingress can be seen in fig. S5A; the absorber layer area diminishes from the edges toward the center, revealing the underlying gold after 800 hours of Humidity Freeze. This is due to the smaller size of the available PO tape that covered only 2 mm of edge margin at the narrowest part, resulting in weaker performance than for the wider PIB tape, which covered the entire edge margin (5 to 6 mm). Wider PO coverage may allow PSCs to sustain the 1000 hours required for the Damp Heat test. The slight increase in V_{OC} (fig. S4B) is due to the slight bandgap increase of the perovskite evidenced by the slight blue shift of the photoluminescence peak (fig. S5B), owing to the possible halide segregation induced by heat and moisture ingress.

PO-encapsulated CsFAMA cells sustained no degradation after 30 cycles of the Humidity Freeze test, again exceeding the IEC 61215:2016 standard. In addition, their PCEs increased by 10 to 20% relative. No moisture ingress was visible in these PSCs after 30 cycles of Humidity Freeze. Figure S4 shows that the J_{SC} and V_{OC} of the cells remained stable and that FF improved by more than 20% relative, which contributed to the observed PCE increase. Without Cs in the perovskite, FAMA cells survived only 10 Humidity Freeze cycles with an average PCE drop of 5%, although they still met the IEC 61215:2016 requirement. The better stability provided by Cs has been verified by GC-MS (see below). Nonetheless, we can conclude that the PO encapsulant is robust in terms of its ability to withstand extreme thermal cycles.

Overall, the excellent stability of the encapsulated PSCs reported here is attributed to the pressure-tight environment provided by the polymer/glass packaging, which effectively suppressed outgassing and thermal decomposition of the PSCs during the accelerated lifetime testing. Both hypotheses are verified by the following GC-MS experiments.

Gas chromatography–mass spectrometry analysis

Automated headspace GC-MS analyses (fig. S6) were carried out on three types of samples in this work: organic perovskite precursor powders, unencapsulated test structures, and encapsulated test cells.

Organic perovskite precursor powders

The organic perovskite precursors MAI, FAI, MABr, and $(FAI)_{0.85}$ + $(MABr)_{0.15}$ in powder form were analyzed with the aim of identifying their major decomposition products. This is because organic components of the perovskites have been reported to be the source of volatile products (*17*, *57*–*59*, *63*, *64*). They were sealed in vials and subsequently annealed on a hotplate at three different conditions: 350°C for 15 min, 140°C for 10 hours, and 85°C for 100 hours (*58*, *64*).

Figure 3 summarizes the peak areas of the chromatograms for the annealed precursor materials. After annealing, CH_3I, HCN, and CH_3Br were found to be the major decomposition products from CH_3NH_3I (MAI), $HC(NH_2)_2I$ (FAI), and CH_3NH_3Br (MABr), respectively (Fig. 3A), as evidenced by the chromatograms obtained and shown in fig. S7, A to C, referencing the mass spectra in fig. S8, A to C. In addition, NH_3 was detected in all precursor materials tested (Fig. 3 and fig. S7E). These findings are in agreement with reports of the same decomposition products in the literature (*57*, *59*, *64*). Decomposition reactions of MAI, FAI, and MABr are shown in Eqs. 1 to 3, respectively, according to results from GC-MS. Other reactions, as shown in Eqs. 4a and 4b, have also been reported to be the decomposition pathways (although reversible) for MAI and MABr, respectively (*61*, *63*). However, our results suggest that the CH_3NH_2 is merely an intermediate product of MA decomposition, as the detected amount was negligible (figs. S7F and S8D) relative to other major decomposition products.

$$CH_3NH_3I(s) \xrightarrow{\Delta} CH_3I(g) + NH_3(g) \quad (1)$$

$$HC(NH_2)_2I(s) \xleftrightarrow{\Delta} HCN(g) + NH_3(g) + HI(g) \quad (2)$$

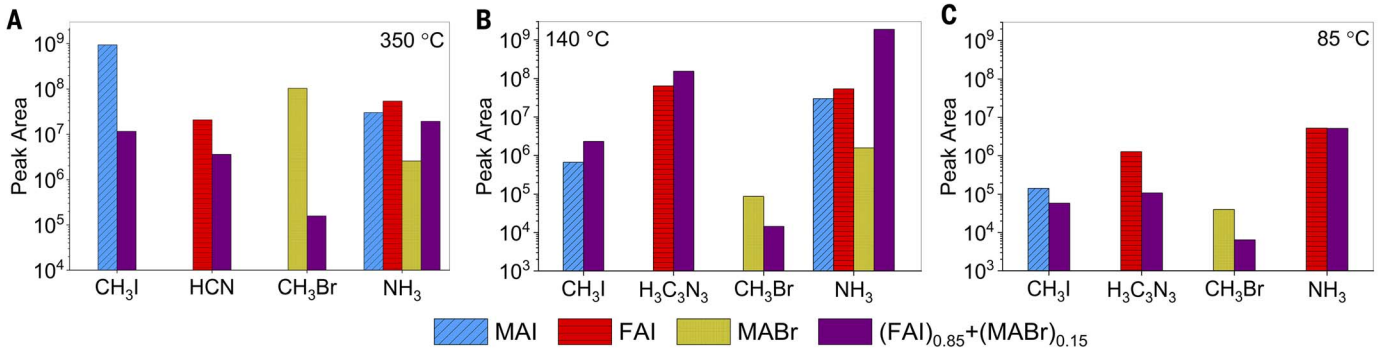


Fig. 3. Signature decomposition products of dry perovskite precursor powders. (**A** to **C**) GC-MS peaks of products after annealing at (A) 350°C for 15 min, (B) 140°C for 10 hours, and (C) 85°C for 100 hours.

$$CH_3NH_3Br(s) \xrightarrow{\Delta} CH_3Br(g)+NH_3(g) \quad (3)$$

$$CH_3NH_3I(s) \xleftrightarrow{\Delta} CH_3NH_2(g)+HI(g) \quad (4a)$$

$$CH_3NH_3Br(s) \xleftrightarrow{\Delta} CH_3NH_2(g)+HBr(g) \quad (4b)$$

For the perovskite precursor FAI annealed at lower temperatures (140°C and 85°C), $H_3C_3N_3$ instead of HCN was found to be the major decomposition product (Fig. 3, B and C, and figs. S9 and S10). Regarding high-temperature (350°C) annealing of the precursor $(FAI)_{0.85}$ + $(MABr)_{0.15}$, the decomposed products of HCN, CH_3Br, and NH_3 detected (Fig. 3A) were expected from the FAI and MABr mixture in 85:15 molar ratio (fig. S7D). However, we also detected CH_3I (Fig. 3A), which is typically a product of MAI decomposition. This could come from the decomposition of a solid solution of $FA_{0.85}MA_{0.15}I_{0.85}Br_{0.15}$ formed by annealing $(FAI)_{0.85}$ + $(MABr)_{0.15}$ at high temperatures (see photos in fig. S11A). This appears to be the case for the 140°C-annealed precursor mixture (fig. S11B) as well, which may be eutectic, given that the melting points of FAI and MABr powders are 242°C and 296°C, respectively.

CH_3I was still detected from $(FAI)_{0.85}$ + $(MABr)_{0.15}$ annealed at the much lower temperature of 85°C (Fig. 3C), despite the lack of solid $FA_{0.85}MA_{0.15}I_{0.85}Br_{0.15}$ solution formed (fig. S11C). Apparently, this CH_3I was mostly produced by the reaction between the CH_3NH_2 and HI as described by Eq. 5 (*63*), where CH_3NH_2 comes from MABr (*65*) (Eq. 4b) and HI comes from FAI (Eq. 2):

$$CH_3NH_2(g)[\text{Eq.4b}] + HI(g)[\text{Eq.2}] \xrightarrow{\Delta} CH_3I(g)+NH_3(g) \quad (5)$$

CH_3CN (acetonitrile) (figs. S7, D and H, S8E, and S9A) is also a product of the decomposition of 350°C- and 140°C-annealed $(FAI)_{0.85}$ + $(MABr)_{0.15}$ detected by the QPlot method. It is believed that CH_3CN is the product of HCN (from FAI; Eq. 2) reacting with CH_3X (from MA-containing precursors; Eq. 1, 3, or 5) through a secondary reaction (Eq. 6):

$$HCN(g)[\text{Eq. 2}] + CH_3X(g)[\text{Eq. 1, 3, or 5}] \rightarrow CH_3CN(g)+HX(g), X = \text{I or Br} \quad (6)$$

The secondary reactions that consume CH_3NH_2 [Eq. 4, which is kinetically favorable (*63*)] and HCN [Eq. 6, which has a negative Gibbs free energy change (−45.3 kJ/mol at 350°C and −46.9 kJ/mol at 140°C) (*66*, *67*)] explain why the CH_3Br and HCN peaks (purple histograms) from $(FAI)_{0.85}$ + $(MABr)_{0.15}$ annealed at 350°C were disproportionally lower than the CH_3Br peak from pure MABr (yellow histograms) and the HCN peak from pure FAI (red histogram) (Fig. 3A). Another signature FA-related decomposition product is $H_3C_3N_3$ (*sym*-triazine) (Eq. 7):

$$3HC(NH_2)_2I(s) \xrightarrow{\Delta} H_3C_3N_3(g)+3NH_3(g)+3HI(g) \quad (7)$$

In addition to $H_3C_3N_3$, other FA-related decomposition products include $H_4C_3N_4$ and CHO(NH_2) (formamide) detected by the Amine method (figs. S7G, S8, I to K, S9C, and S10A). $H_4C_3N_4$ detected from the 350°C-annealed precursor could be the product of a reaction between $H_3C_3N_3$ and NH_3 at high temperature. $CH(N_2H_3)$ (formamidine), as reported in (*64*), was not directly detected here, but CHO(NH_2), possibly in its oxidized form, was detected, although the source of oxygen is yet to be ascertained.

Another minor product, CH_3(NH)CHO (methylformamide), was detected specifically from $(FAI)_{0.85}$ + $(MABr)_{0.15}$ (figs. S7G, S8L, S9C, and S10C). This could be formed by CHO(NH_2) reacting with CH_3I, with HI also being a product. Alternatively, CH_3(NH)CHO could be formed by CHO(NH_2) reacting with CH_3NH_2, with NH_3 also being a product (*68*). Methylformamide, more so than CH_3CN, is an interaction product of low temperature–annealed MA and FA mixed powders (fig. S10A).

HI could be detected by the Amine method (fig. S9C) and more reliably by the direct MS analysis method from annealed MAI, FAI, and $(FAI)_{0.85}$ + $(MABr)_{0.15}$ (fig. S12), as predicted by Eqs. 2 and 4a. Likewise, HBr could be detected by the direct MS analysis method from annealed MABr (fig. S13), as predicted by Eq. 3.

For 350°C- and 140°C-annealed precursors, the amount of NH_3 produced was similar from all precursors except MABr, which suggests that MABr is less prone to thermal decomposition. This is because Br^- is a stronger conjugated base of hydrogen halide than I^- (*61*) and has a much higher oxidation potential (*69*), thus favoring the CH_3NH_2/HBr decomposition pathway (Eq. 4b) over the detrimental CH_3Br/NH_3 route (Eq. 3).

From low temperature–annealed MAI and MABr, no NH_3 was detected, possibly because of the reaction between CH_3X (X = I or Br) and NH_3 at low temperature producing ammonia salt, as reported in (*61*). The lack of contribution from MABr to mixed $(FAI)_{0.85}$ + $(MABr)_{0.15}$ in terms of NH_3 product explains the almost identical NH_3 peaks from pure FAI and mixed $(FAI)_{0.85}$ + $(MABr)_{0.15}$ (Fig. 3C).

GC-MS experiments were also performed on CsI and $(CsI)_{0.05}$ + $(FAI)_{0.8}$ + $(MABr)_{0.15}$ powders. The chromatograms (fig. S7H) were similar to those of mixed FAMA precursors without Cs, which suggests no outgassing from CsI (fig. S7I).

These findings illustrate that the mixing of MA with FA perovskite reduces stability, producing decomposition products that otherwise would not be predicted in FA-only perovskites. Although the decomposition of FAI as represented by Eq. 2 is reversible (*64*), when FAI is mixed with MA-containing precursors, the decomposition products participate in the secondary reaction (e.g., Eq. 5), which is irreversible (*61*).

Figure 4 shows the types of unencapsulated test structures and encapsulated test cells analyzed in this work, each designated with a label with prefix "Con." For the GC-MS analyses, these samples were sealed in the vials and

annealed at 85°C for 100 hours to encourage outgassing before GC-MS measurements.

Unencapsulated test structures

Figure 5A compares the GC results of 85°C-annealed unencapsulated test structures Con. 1 (FAMA/PTAA), Con. 2 (PTAA), and Con. 3 (FAMA), all of which were fabricated on FTO/c-TiO_2/mp-TiO_2 glass substrate and covered by Au. After 100 hours, no outgassing could be detected from PTAA thin film (Con. 2) by either the Amine method (Fig. 5A, orange curve) or the QPlot method (fig. S14). This is an encouraging demonstration of the good stability of PTAA.

Similar to the results for 85°C-annealed mixed perovskite precursor powders, the signature decomposition products CH_3I (Fig. 5B), CH_3Br (Fig. 5C), and NH_3 (Fig. 5D) could be detected in unencapsulated Con. 1 test cells using single ion monitoring (SIM) mode by either method. $CHO(NH_2)$, $CH_3(NH)CHO$ (fig. S15F), and a small amount of CH_3NH_2 (intermediate decomposition product) were also detected (see fig. S15G). HCN and $H_3C_3N_3$ were not detected; the absence of HCN as a decomposition product of unencapsulated cells annealed at low temperature agrees with results reported by Ma *et al.* (*57*), who observed HCN release only at temperatures above 285°C and 350°C for FAI and $FAPbI_3$, respectively.

More important, CsFAMA cells outgassed one-fifth as much decomposition product (compare the orange and blue histograms for Con. 1 samples in Fig. 5, B to D) as their FAMA counterparts, indicating better thermal stability of Cs-containing cells.

Encapsulated test cells

In encapsulated cells, the intensities of CH_3I (Fig. 5B), CH_3Br (Fig. 5C), and NH_3 (Fig. 5D) decomposition products were reduced. In addition, CH_3NH_2, HI, $CHO(NH_2)$, and $CH_3(NH)CHO$ could not be detected (fig. S15). PO (Con. 1/PO) was more effective than PIB (Con. 1/PIB) in suppressing CH_3I and CH_3Br outgassing, especially for FAMA cells. Although PO in itself was more permeable to NH_3 than PIB (probably due to the better absorption of NH_3 by PIB during the test), the application of cover glass eliminated NH_3 outgassing in any case (Fig. 5D).

For edge-sealed PSCs, a syringe was used to sample the gas within the cavity between the PSC and the cover glass to look for the major decomposition products. This is a nonintrusive way to analyze chemical decomposition processes without having to disassemble the packaged PSC. Relative to unencapsulated cells, the amounts of CH_3Br and CH_3I detected from edge-sealed cells were also reduced (Fig. 5, B and C, and fig. S15, A and B), showing the effectiveness of the edge sealing.

C_3H_6 was detected from encapsulated structures only (fig. S15, A and B); this C_3H_6 is believed to originate from PIB- or PO-based polymer outgassing and is particularly prominent in manually sampled GC (because of direct contact between PIB and the syringe during manual sampling). Most important, C_3H_6 was successfully distinguished from other perovskite-related signals with similar MS patterns, such as CH_3CN, owing to the high specificity of GC-MS.

The application of cover glass on encapsulated cells successfully suppressed CH_3Br and NH_3 outgassing for either CsFAMA or FAMA cells using any type of encapsulation: blanket or edge-sealed CsFAMA using either PIB or PO. It can be concluded that neither PO nor PIB can completely suppress perovskite outgassing without the use of cover glass. Nonetheless, the polymers function effectively as moisture

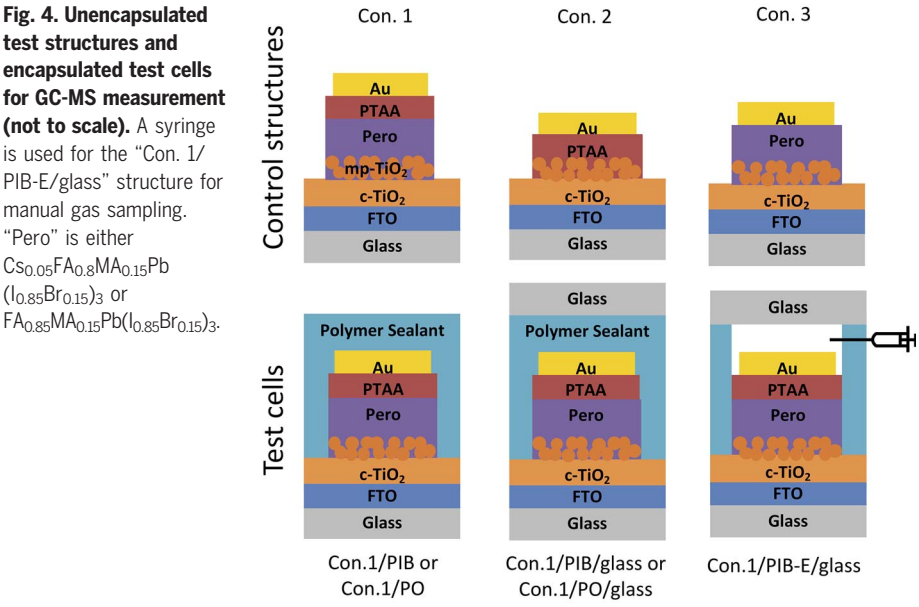


Fig. 4. Unencapsulated test structures and encapsulated test cells for GC-MS measurement (not to scale). A syringe is used for the "Con. 1/PIB-E/glass" structure for manual gas sampling. "Pero" is either $Cs_{0.05}FA_{0.8}MA_{0.15}Pb(I_{0.85}Br_{0.15})_3$ or $FA_{0.85}MA_{0.15}Pb(I_{0.85}Br_{0.15})_3$.

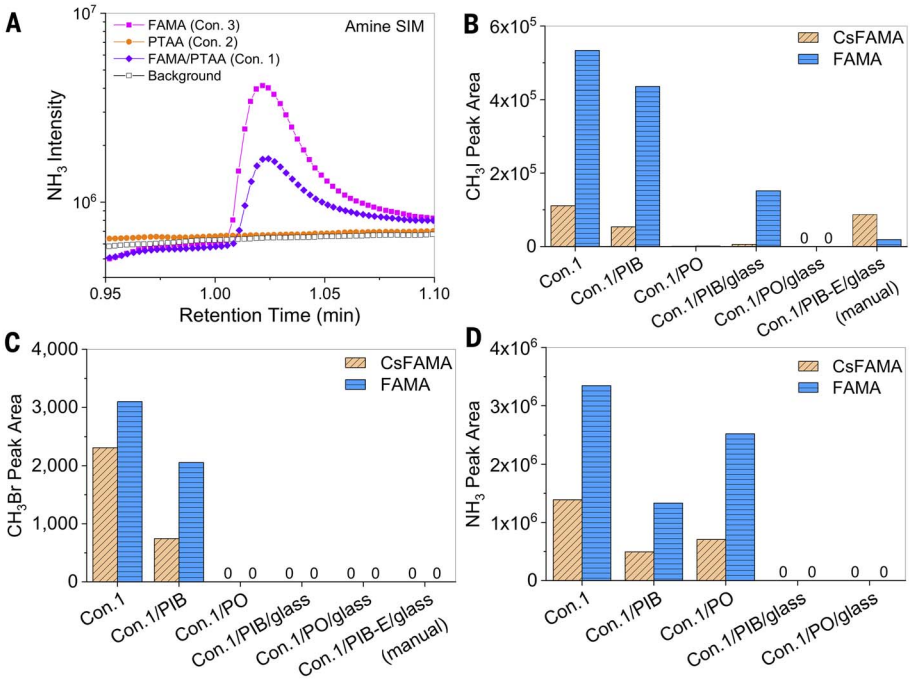


Fig. 5. Signature decomposition products of unencapsulated test structures and encapsulated test cells. (**A**) GC chromatograms for unencapsulated test structures Con. 3 (FAMA), Con. 2 (PTAA), and Con. 1 (FAMA/PTAA) annealed for 100 hours at 85°C. (**B** to **D**) SIM chromatographic peak areas of the signature decomposition products of CH_3I (B), CH_3Br (C), and NH_3 (D) for PSCs with or without packaging, annealed for 100 hours at 85°C.

barriers and as strong adhesives between the PSC and the cover glass. They also have low elastic modulus (0.5 MPa for PIB; 20 MPa for PO) (*70*) and therefore maintain the encapsulation integrity during the harsh conditions simulated by the Humidity Freeze test (*71*).

The polymer/glass blanket encapsulation is an effective scheme providing a pressure-tight environment for the PSC. This causes any decomposition reaction to reach equilibrium quickly before damage occurs to the PSC, as described by Eq. 8. ΔG is the free energy change during the decomposition reaction, $\Delta G°$ is the free energy change when the reactants and products are in their standard states, and Q is the ratio of the product concentration to the reactant concentration. As the concentration of products increases in a pressure-tight environment, Q increases and ΔG tends toward zero, indicating that the reaction reaches equilibrium. The variation of the temperature could even drive ΔG to positive (i.e., unspontaneous) reaction. Sealing the PSC in a pressure-tight environment effectively suppresses the decomposition reaction even when it is not fully reversible, as is the case for MA or FA/MA mixed multi-cation perovskites.

$$\Delta G = \Delta G° + RT \ln Q \qquad (8)$$

Finally, preliminary evaluation of the blanket encapsulation (PIB and PO) scheme by maximum power point tracking (MPPT) with the purpose of evaluating the temperature threshold under constant illumination was carried out on CsFAMA PSCs. The results (fig. S16) show that blanket-encapsulated cells are more stable than edge-sealed cells. Temperatures above 35°C trigger degradations. It is expected that cells with architectures that are less sensitive to ultraviolet light will have better stability.

Conclusion

Our work shows that organic-inorganic PSCs encapsulated with simple and low-cost polyisobutylene- or polyolefin-based polymer-glass combinations have excellent durability exceeding the requirements of the IEC 61215:2016 Damp Heat and Humidity Freeze tests. These results are especially noteworthy because the PSCs used in this study contain MA cations, which are known to have lower thermal stability. Yet these cells survived the Humidity Freeze test, which exerts the harshest humidity and thermal stress. Using GC-MS, we have identified the decomposition products of single, mixed-cation, and mixed-halide perovskite precursor materials as well as completed PSCs with high specificity. On the basis of the identified signature decomposition products such as CH_3I, CH_3Br, and NH_3, we propose decomposition pathways for MAI, FAI, MABr, mixed-cation, and mixed-halide perovskite precursors including their secondary decomposition reactions. We then found that although decomposition pathways are similar for complete PSCs, decomposition is effectively suppressed by the use of polymer–glass blanket encapsulation, which is a hermetic encapsulation scheme creating a pressure-tight environment and preventing the escape of decomposition products. The developments in this work contributes valuable knowledge and advances for stabilizing PSCs, increasing their commercialization prospects.

Methods summary

Details of materials and processes for PSC fabrication, packaging, and characterization can be found in the supplementary materials.

Similar to previous reports, electrical feedthrough was achieved without the use of metal but instead via the FTO layer, ensuring hermeticity (*25*). However, additional edge cleaning steps were applied in this work to ensure good adhesion between the encapsulant and the c-TiO_2 layer.

Automated headspace GC-MS analysis (fig. S6) was performed to investigate the temperature-induced decomposition of PSCs and the effectiveness of different packaging schemes. A sample under analysis was first sealed in a glass vial inside a N_2-filled glovebox before being annealed to generate gaseous substances from the thermal decomposition of the sample inside the vial. After cooling to room temperature, the gas inside the vial was sampled and injected onto the column by the automated headspace GC-MS system. In addition to normal mass-to-charge ratio (m/z) scan mass spectra, SIM was used as a supplementary technique to quantify weaker signals (e.g., for encapsulated cells with very low amounts of CH_3I, HCN, CH_3Br, and NH_3 outgassing). The intensity of each peak was quantified by the peak area using the software. Two columns, a Restek QPLOT column (QPlot method) and a base-deactivated Restek VolatileAmine column (Amine method), were used for GC-MS analysis in this work with different levels of sensitivity. To further improve the specificity of GC-MS results, the identification of key decomposition products was verified by GC-MS measurement of the same standard gas samples (fig. S17, A to D). We used direct MS analysis to detect the highly reactive HI and HBr gases (fig. S17E); these gases could not be reliably detected by GS-MS because of their reaction with the stationary phase of the columns (*72*). More details of the GC-MS methods can be found in the supplementary materials.

REFERENCES AND NOTES

1. M. A. Green, A. Ho-Baillie, H. J. Snaith, The emergence of perovskite solar cells. *Nat. Photonics* **8**, 506–514 (2014). doi: 10.1038/nphoton.2014.134
2. M. A. Green, A. Ho-Baillie, Perovskite Solar Cells: The Birth of a New Era in Photovoltaics. *ACS Energy Lett.* **2**, 822–830 (2017). doi: 10.1021/acsenergylett.7b00137
3. M. A. Green *et al*., Solar cell efficiency tables (version 54). *Prog. Photovolt. Res. Appl.* **28**, 3–15 (2020). doi: 10.1002/pip.3228
4. National Renewable Energy Laboratory, Efficiency Chart (2019); www.nrel.gov/pv/cell-efficiency.html.
5. N. L. Chang *et al*., A manufacturing cost estimation method with uncertainty analysis and its application to perovskite on glass photovoltaic modules. *Prog. Photovolt. Res. Appl.* **25**, 390–405 (2017). doi: 10.1002/pip.2871
6. J. H. Noh, S. H. Im, J. H. Heo, T. N. Mandal, S. I. Seok, Chemical management for colorful, efficient, and stable inorganic-organic hybrid nanostructured solar cells. *Nano Lett.* **13**, 1764–1769 (2013). doi: 10.1021/nl400349b; pmid: 23517331
7. S. N. Habisreutinger *et al*., Carbon nanotube/polymer composites as a highly stable hole collection layer in perovskite solar cells. *Nano Lett.* **14**, 5561–5568 (2014). doi: 10.1021/nl501982b; pmid: 25226226
8. G. Niu *et al*., Study on the stability of $CH_3NH_3PbI_3$ films and the effect of post-modification by aluminum oxide in all-solid-state hybrid solar cells. *J. Mater. Chem. A* **2**, 705–710 (2014). doi: 10.1039/C3TA13606J
9. J. A. Christians, P. A. Miranda Herrera, P. V. Kamat, Transformation of the excited state and photovoltaic efficiency of $CH_3NH_3PbI_3$ perovskite upon controlled exposure to humidified air. *J. Am. Chem. Soc.* **137**, 1530–1538 (2015). doi: 10.1021/ja511132a; pmid: 25590693
10. M. I. Asghar, J. Zhang, H. Wang, P. D. Lund, Device stability of perovskite solar cells – A review. *Renew. Sustain. Energy Rev.* **77**, 131–146 (2017). doi: 10.1016/j.rser.2017.04.003
11. J. Huang, S. Tan, P. Lund, H. Zhou, Impact of H_2O on organic-inorganic hybrid perovskite solar cells. *Energy Environ. Sci.* **10**, 2284–2311 (2017). doi: 10.1039/C7EE01674C
12. P. Wang *et al*., Solution-Processable Perovskite Solar Cells toward Commercialization—Progress and Challenges. *Adv. Funct. Mater.* **29**, 1807661 (2019). doi: 10.1002/adfm.201807661
13. A. Dualeh *et al*., Effect of annealing temperature on film morphology of organic-inorganic hybrid pervoskite solid-state solar cells. *Adv. Funct. Mater.* **24**, 3250–3258 (2014). doi: 10.1002/adfm.201304022
14. B. Conings *et al*., Intrinsic Thermal Instability of Methylammonium Lead Trihalide Perovskite. *Adv. Energy Mater.* **5**, 1500477 (2015). doi: 10.1002/aenm.201500477
15. B. Philippe *et al*., Chemical and Electronic Structure Characterization of Lead Halide Perovskites and Stability Behavior under Different Exposures - a Photoelectron Spectroscopy Investigation. *Chem. Mater.* **27**, 1720–1731 (2015). doi: 10.1021/acs.chemmater.5b00348
16. N. K. Kim *et al*., Investigation of Thermally Induced Degradation in $CH_3NH_3PbI_3$ Perovskite Solar Cells using In-situ Synchrotron Radiation Analysis. *Sci. Rep.* **7**, 4645 (2017). doi: 10.1038/s41598-017-04690-w; pmid: 28680138
17. C. C. Boyd, R. Cheacharoen, T. Leijtens, M. D. McGehee, Understanding Degradation Mechanisms and Improving Stability of Perovskite Photovoltaics. *Chem. Rev.* **119**, 3418–3451 (2019). doi: 10.1021/acs.chemrev.8b00336; pmid: 30444609
18. T. Leijtens *et al*., Overcoming ultraviolet light instability of sensitized TiO_2 with meso-superstructured organometal tri-halide perovskite solar cells. *Nat. Commun.* **4**, 2885 (2013). doi: 10.1038/ncomms3885; pmid: 24301460
19. S. Ito, S. Tanaka, K. Manabe, H. Nishino, Effects of Surface Blocking Layer of Sb_2S_3 on Nanocrystalline TiO_2 for $CH_3NH_3PbI_3$ Perovskite Solar Cells. *J. Phys. Chem. C* **118**, 16995 (2014). doi: 10.1021/jp500449z
20. J. Ahmad, K. Bazaka, L. J. Anderson, R. D. White, M. V. Jacob, Materials and methods for encapsulation of OPV: A review. *Renew. Sustain. Energy Rev.* **27**, 104–117 (2013). doi: 10.1016/j.rser.2013.06.027
21. D. Yu, Y. Q. Yang, Z. Chen, Y. Tao, Y. F. Liu, Recent progress on thin-film encapsulation technologies for organic electronic devices. *Opt. Commun.* **362**, 43–49 (2016). doi: 10.1016/j.optcom.2015.08.021
22. A. Uddin, M. Upama, H. Yi, L. Duan, Encapsulation of Organic and Perovskite Solar Cells: A Review. *Coatings* **9**, 65 (2019). doi: 10.3390/coatings9020065
23. M. D. Kempe, A. A. Dameron, M. O. Reese, Evaluation of moisture ingress from the perimeter of photovoltaic modules. *Prog. Photovolt. Res. Appl.* **22**, 1159–1171 (2014). doi: 10.1002/pip.2374
24. F. J. Pern, R. Noufi, paper presented at SPIE Optics + Photonics, San Diego, CA, 16 October 2012.
25. L. Shi *et al*., Accelerated Lifetime Testing of Organic-Inorganic Perovskite Solar Cells Encapsulated by Polyisobutylene. *ACS Appl. Mater. Interfaces* **9**, 25073–25081 (2017). doi: 10.1021/acsami.7b07625; pmid: 28700216

26. F. Matteocci *et al*., Encapsulation for long-term stability enhancement of perovskite solar cells. *Nano Energy* **30**, 162–172 (2016). doi: 10.1016/j.nanoen.2016.09.041
27. Y. Han *et al*., Degradation observations of encapsulated planar $CH_3NH_3PbI_3$ perovskite solar cells at high temperatures and humidity. *J. Mater. Chem. A* **3**, 8139–8147 (2015). doi: 10.1039/C5TA00358J
28. H. C. Weerasinghe *et al*., New barrier encapsulation and lifetime assessment of printed organic photovoltaic modules. *Sol. Energy Mater. Sol. Cells* **155**, 108–116 (2016). doi: 10.1016/j.solmat.2016.04.051
29. Q. Dong *et al*., Encapsulation of Perovskite Solar Cells for High Humidity Conditions. *ChemSusChem* **9**, 2597–2603 (2016). doi: 10.1002/cssc.201601090; pmid: 27504719
30. J. Kim *et al*., An effective method of predicting perovskite solar cell lifetime–Case study on planar $CH_3NH_3PbI_3$ and $HC(NH_2)_2PbI_3$ perovskite solar cells and hole transfer materials of spiro-OMeTAD and PTAA. *Sol. Energy Mater. Sol. Cells* **162**, 41–46 (2017). doi: 10.1016/j.solmat.2016.12.043
31. G. E. Eperon *et al*., Formamidinium lead trihalide: A broadly tunable perovskite for efficient planar heterojunction solar cells. *Energy Environ. Sci.* **7**, 982 (2014). doi: 10.1039/c3ee43822h
32. W. Tan, A. R. Bowring, A. C. Meng, M. D. McGehee, P. C. McIntyre, Thermal Stability of Mixed Cation Metal Halide Perovskites in Air. *ACS Appl. Mater. Interfaces* **10**, 5485–5491 (2018). doi: 10.1021/acsami.7b15263; pmid: 29328620
33. C. Yi *et al*., Entropic stabilization of mixed A-cation ABX_3 metal halide perovskites for high performance perovskite solar cells. *Energy Environ. Sci.* **9**, 656–662 (2016). doi: 10.1039/C5EE03255E
34. M. Saliba *et al*., Incorporation of rubidium cations into perovskite solar cells improves photovoltaic performance. *Science* **354**, 206–209 (2016). doi: 10.1126/science.aah5557; pmid: 27708053
35. G. Grancini *et al*., One-Year stable perovskite solar cells by 2D/3D interface engineering. *Nat. Commun.* **8**, 15684 (2017). doi: 10.1038/ncomms15684; pmid: 28569749
36. N. Li *et al*., Mixed Cation $FA_xPEA_{1-x}PbI_3$ with Enhanced Phase and Ambient Stability toward High-Performance Perovskite Solar Cells. *Adv. Energy Mater.* **7**, 1601307 (2017). doi: 10.1002/aenm.201601307
37. Z. Wang *et al*., Efficient ambient-air-stable solar cells with 2D-3D heterostructured butylammonium-caesium-formamidinium lead halide perovskites. *Nat. Energy* **2**, 17135 (2017). doi: 10.1038/nenergy.2017.135
38. H. Tsai *et al*., High-efficiency two-dimensional Ruddlesden-Popper perovskite solar cells. *Nature* **536**, 312–316 (2016). doi: 10.1038/nature18306; pmid: 27383783
39. A. R. bin M. Yusoff, M. K. Nazeeruddin, Low-Dimensional Perovskites: From Synthesis to Stability in Perovskite Solar Cells. *Adv. Energy Mater.* **8**, 1702073 (2018). doi: 10.1002/aenm.201702073
40. Y. C. Kim *et al*., Engineering interface structures between lead halide perovskite and copper phthalocyanine for efficient and stable perovskite solar cells. *Energy Environ. Sci.* **10**, 2109–2116 (2017). doi: 10.1039/C7EE01931A
41. Y. Chen *et al*., Design of an Inorganic Mesoporous Hole-Transporting Layer for Highly Efficient and Stable Inverted Perovskite Solar Cells. *Adv. Mater.* **30**, 1805660 (2018). doi: 10.1002/adma.201805660
42. N. Arora *et al*., Perovskite solar cells with CuSCN hole extraction layers yield stabilized efficiencies greater than 20. *Science* **358**, 768–771 (2017). doi: 10.1126/science.aam5655; pmid: 28971968
43. E. Bi *et al*., Diffusion engineering of ions and charge carriers for stable efficient perovskite solar cells. *Nat. Commun.* **8**, 15330 (2017). doi: 10.1038/ncomms15330; pmid: 28604673
44. H. Tan *et al*., Efficient and stable solution-processed planar perovskite solar cells via contact passivation. *Science* **355**, 722–726 (2017). doi: 10.1126/science.aai9081; pmid: 28154242
45. M. Saliba *et al*., Cesium-containing triple cation perovskite solar cells: Improved stability, reproducibility and high efficiency. *Energy Environ. Sci.* **9**, 1989–1997 (2016). doi: 10.1039/C5EE03874J; pmid: 27478500
46. K. A. Bush *et al*., 23.6%-efficient monolithic perovskite/silicon tandem solar cells with improved stability. *Nat. Energy* **2**, 17009 (2017). doi: 10.1038/nenergy.2017.9
47. S. Seo, S. Jeong, C. Bae, N. G. Park, H. Shin, Perovskite Solar Cells with Inorganic Electron- and Hole-Transport Layers Exhibiting Long-Term (≈500 h) Stability at 85°C under Continuous 1 Sun Illumination in Ambient Air. *Adv. Mater.* **30**, e1801010 (2018). doi: 10.1002/adma.201801010; pmid: 29786887
48. K. A. Bush *et al*., Thermal and Environmental Stability of Semi-Transparent Perovskite Solar Cells for Tandems Enabled by a Solution-Processed Nanoparticle Buffer Layer and Sputtered ITO Electrode. *Adv. Mater.* **28**, 3937–3943 (2016). doi: 10.1002/adma.201505279; pmid: 26880196
49. C. C. Boyd *et al*., Barrier Design to Prevent Metal-Induced Degradation and Improve Thermal Stability in Perovskite Solar Cells. *ACS Energy Lett.* **3**, 1772–1778 (2018). doi: 10.1021/acsenergylett.8b00926
50. Y. I. Lee *et al*., A Low-Temperature Thin-Film Encapsulation for Enhanced Stability of a Highly Efficient Perovskite Solar Cell. *Adv. Energy Mater.* **8**, 1701928 (2018). doi: 10.1002/aenm.201701928
51. K. O. Brinkmann *et al*., Suppressed decomposition of organometal halide perovskites by impermeable electron-extraction layers in inverted solar cells. *Nat. Commun.* **8**, 13938 (2017). doi: 10.1038/ncomms13938; pmid: 28067308
52. R. Cheacharoen *et al*., Encapsulating perovskite solar cells to withstand damp heat and thermal cycling. *Sustain. Energy Fuels* **2**, 2398–2406 (2018). doi: 10.1039/C8SE00250A
53. Z. Li *et al*., Stabilizing Perovskite Structures by Tuning Tolerance Factor: Formation of Formamidinium and Cesium Lead Iodide Solid-State Alloys. *Chem. Mater.* **28**, 284–292 (2016). doi: 10.1021/acs.chemmater.5b04107
54. A. Dualeh, P. Gao, S. I. Seok, M. K. Nazeeruddin, M. Gratzel, Thermal Behavior of Methylammonium Lead-trihalide Perovskite Photovoltaic Light Harvesters. *Chem. Mater.* **26**, 6160–6164 (2014). doi: 10.1021/cm502468k
55. D. P. Nenon *et al*., Structural and chemical evolution of methylammonium lead halide perovskites during thermal processing from solution. *Energy Environ. Sci.* **9**, 2072–2082 (2016). doi: 10.1039/C6EE01047D
56. A. E. Williams *et al*., Perovskite processing for photovoltaics: A spectro-thermal evaluation. *J. Mater. Chem. A* **2**, 19338–19346 (2014). doi: 10.1039/C4TA04725G
57. L. Ma *et al*., Temperature-dependent thermal decomposition pathway of organic-inorganic halide perovskite materials. *Chem. Mater.* **31**, 8515–8522 (2019). doi: 10.1021/acs.chemmater.9b03190
58. E. J. Juarez-Perez, Z. Hawash, S. R. Raga, L. K. Ono, Y. Qi, Thermal degradation of $CH_3NH_3PbI_3$ perovskite into NH_3 and CH_3I gases observed by coupled thermogravimetry-mass spectrometry analysis. *Energy Environ. Sci.* **9**, 3406–3410 (2016). doi: 10.1039/C6EE02016J
59. E. J. Juarez-Perez *et al*., Photodecomposition and thermal decomposition in methylammonium halide lead perovskites and inferred design principles to increase photovoltaic device stability. *J. Mater. Chem. A* **6**, 9604–9612 (2018). doi: 10.1039/C8TA03501F
60. Z. Fan *et al*., Layer-by-Layer Degradation of Methylammonium Lead Tri-iodide Perovskite Microplates. *Joule* **1**, 548–562 (2017). doi: 10.1016/j.joule.2017.08.005
61. E. J. Juarez-Perez, L. K. Ono, I. Uriarte, E. J. Cocinero, Y. Qi, Degradation Mechanism and Relative Stability of Methylammonium Halide Based Perovskites Analyzed on the Basis of Acid-Base Theory. *ACS Appl. Mater. Interfaces* **11**, 12586–12593 (2019). doi: 10.1021/acsami.9b02374; pmid: 30848116
62. International Electrotechnical Commission, *Terrestrial photovoltaic (PV) modules - Design qualification and type approval - Part 1: Test requirements*; https://webstore.iec.ch/publication/24312.
63. A. Latini, G. Gigli, A. Ciccioli, A study on the nature of the thermal decomposition of methylammonium lead iodide perovskite, $CH_3NH_3PbI_3$: An attempt to rationalise contradictory experimental results. *Sustain. Energy Fuels* **1**, 1351–1357 (2017). doi: 10.1039/C7SE00114B
64. E. J. Juarez-Perez, L. K. Ono, Y. Qi, Thermal degradation of formamidinium based lead halide perovskites into sym-triazine and hydrogen cyanide observed by coupled thermogravimetry - mass spectrometry analysis. *J. Mater. Chem. A* **7**, 16912–16919 (2019). doi: 10.1039/C9TA06058H
65. I. L. Ivanov *et al*., Thermodynamics of formation of hybrid perovskite-type methylammonium lead halides. *J. Chem. Thermodyn.* **116**, 253–258 (2018). doi: 10.1016/j.jct.2017.09.026
66. J. Speight, *Lange's Handbook of Chemistry* (McGraw-Hill, ed. 16, 2005); www.accessengineeringlibrary.com/content/book/9780071432207/toc-chapter/chapter1/section/section83.
67. NIST Chemistry WebBook; https://webbook.nist.gov/chemistry/.
68. H. Bipp, H. Kieczka, Formamides. In *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley-VCH, ed. 7, 2011), vol. 16.
69. A. F. Akbulatov *et al*., Light or Heat: What Is Killing Lead Halide Perovskites under Solar Cell Operation Conditions? *J. Phys. Chem. Lett.* **11**, 333–339 (2020). doi: 10.1021/acs.jpclett.9b03308; pmid: 31838849
70. B. A. Turowec, E. R. Gillies, Synthesis, properties and degradation of polyisobutylene–polyester graft copolymers. *Polym. Int.* **66**, 42–51 (2017). doi: 10.1002/pi.5250
71. R. Cheacharoen *et al*., Design and understanding of encapsulated perovskite solar cells to withstand temperature cycling. *Energy Environ. Sci.* **11**, 144–150 (2018). doi: 10.1039/C7EE02564E
72. Gas Chromatography Problem Solving and Troubleshooting. *J. Chromatogr. Sci.* **37**, 313 (1999). doi: 10.1093/chromsci/37.8.313

ACKNOWLEDGMENTS

We thank X. Deng for assistance in photoluminescence measurement, J. Zheng for assistance in proofreading the manuscript and providing valuable suggestions, and N. Proschogo (University of Sydney) for assistance in evaluating initial GC-MS techniques. **Funding:** The Australian Centre for Advanced Photovoltaics (ACAP) encompasses the Australian-based activities of the Australia-U.S. Institute for Advanced Photovoltaics (AUSIAPV) and is supported by the Australian Government through the Australian Renewable Energy Agency (ARENA). L.S. also thanks ARENA for scholarship funding support. **Author contributions:** L.S. and A.W.Y.H.-B. conceived and designed all the experimental work. L.S., T.L.Y., M.Z., D.S.L., J.B., and J.K. fabricated samples (perovskite thin-films and devices). L.S. encapsulated the samples and performed measurements. M.Z. assisted in producing the figures. M.P.B. contributed to the GC-MS methodology and assisted the measurement. N.T. and D.R.M. contributed to initiating the GC-MS idea and commenced initial measurements. L.H. and T.W. contributed to the discussion of pressure-tight sealing and decomposition reactions. The manuscript was written by L.S., M.P.B., and A.W.Y.H.-B. All authors contributed to the discussion of the data, proofread of the manuscript. The overall project was supervised by M.A.G., S.H., and A.W.Y.H.-B. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/eaba2412/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 to S7
References (*73*–*78*)

16 November 2019; accepted 8 May 2020
Published online 21 May 2020
10.1126/science.aba2412

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Manipulating synthetic optogenetic odors reveals the coding logic of olfactory perception

Edmund Chong*, Monica Moroni*, Christopher Wilson, Shy Shoham, Stefano Panzeri, Dmitry Rinberg†

INTRODUCTION: Advances in monitoring brain activity at large and fine scales have revealed tremendous complexity in how the brain responds to, and represents, the external world. Although many features in brain activity patterns (which brain cells fire and when) are found to correlate with changes in the external sensory world, it is not yet known which activity features are consequential for perception and how they are combined to generate percepts. Some studies have shown that many of these correlated changes in activity may be redundant or even epiphenomenal.

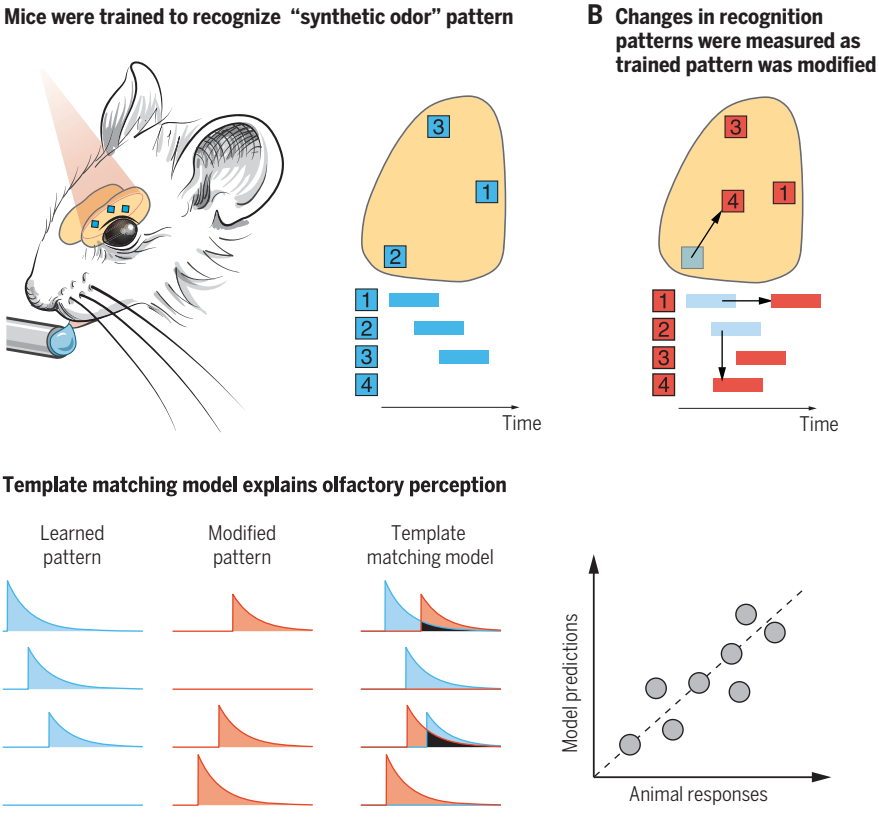
RATIONALE: To address how brain activity generates perception, we directly and systematically manipulated neural activity in the mouse olfactory system while measuring perceptual responses. Mouse olfaction is an attractive model system because the relevant brain circuitry has already been carefully mapped out and is accessible for direct manipulation. We used genetically engineered mice in which brain cells can be activated simply by shining light on them—a technique known as optogenetics. Optogenetics allowed us to directly generate and manipulate brain activity in a precise and parametric manner. We first trained mice to recognize light-driven activity patterns in the olfactory system, or "synthetic odors." Subsequently, we measured how recognition changed as we systematically manipulated learned activity patterns. Some manipulations led to larger changes in recognition than others, and the degree of change reflected the importance of each manipulated feature to perception. By the additional manipulation of multiple features simultaneously, we could precisely quantify how individual features combined to produce perception.

RESULTS: The perceptual responses of mice not only depended on which groups of cells were activated, but also on their activation latencies, i.e., temporal sequences akin to timed notes in a melody. Critically, the most perceptually relevant activation latencies were defined relative to other cells in a sequence and not to brain or body rhythms (e.g., animal sniffing) as previously hypothesized from observational studies. Moreover, earlier-activated cells in the sequence had a larger effect on behavioral responses; modifying later cells in the sequence had small effects. To account for all results, we formulated a simple computational model based on template matching, in which new activity sequences are compared with learned sequences or templates. The model weighs relative timing within each sequence and also accounts for the greater importance of earlier-activated cells. Based on our model, the degree of mismatch between the new sequence and learned template predicts the extent to which recognition should degrade as neural activity changes across many different manipulations.

ON OUR WEBSITE

Read the full article at https://dx.doi.org/10.1126/science.aba2357

CONCLUSION: We developed an experimental and theoretical framework to map a broad space of precisely and systematically manipulated brain activity patterns to behavioral responses. Using this framework, we uncovered key computations made by the olfactory system on neural activity to generate percepts and derived a systematic model of olfactory processing directly relevant for perception. Our framework forms a powerful, general approach for causally testing the links between brain activity and perception or behavior. This framework is especially pertinent given the continued development of advanced tools for manipulating brain activity at fine scales across various brain regions. ■



Probing olfactory perception with synthetic odors. (**A**) We trained mice to recognize synthetic odor patterns: artificially stimulated neural activity in the olfactory bulb. Patterns were defined in space (top right) and time (bottom right). (**B**) Perceptual responses were measured across systematic modifications of trained patterns. (**C**) Template-matching model of pattern activity (left) accounts for perceptual responses (right).

*These authors contributed equally to this work.
†Corresponding author. Email: rinberg@nyu.edu
The list of author affiliations is available in the full article online.
Cite this article as E. Chong *et al.*, *Science* **368**, eaba2357 (2020). DOI: 10.1126/science.aba2357

RESEARCH ARTICLE

NEUROSCIENCE

Manipulating synthetic optogenetic odors reveals the coding logic of olfactory perception

Edmund Chong[1]*, Monica Moroni[2,3]*, Christopher Wilson[1], Shy Shoham[1,4,5,6], Stefano Panzeri[2], Dmitry Rinberg[1,4]†

How does neural activity generate perception? Finding the combinations of spatial or temporal activity features (such as neuron identity or latency) that are consequential for perception remains challenging. We trained mice to recognize synthetic odors constructed from parametrically defined patterns of optogenetic activation, then measured perceptual changes during extensive and controlled perturbations across spatiotemporal dimensions. We modeled recognition as the matching of patterns to learned templates. The templates that best predicted recognition were sequences of spatially identified units, ordered by latencies relative to each other (with minimal effects of sniff). Within templates, individual units contributed additively, with larger contributions from earlier-activated units. Our synthetic approach reveals the fundamental logic of the olfactory code and provides a general framework for testing links between sensory activity and perception.

A familiar object evokes a complex pattern of activity in the brain. However, it is possible that only a structured subset of this activity, representing critical combinations of sensory attributes, is essential for recognition. A key challenge is to identify the essential components of neural activity that induce the percept. This activity may consist of multiple spatial or temporal features, such as which cells respond and when they respond relative to stimulus onset or each other. Do individual features contribute differentially to the formation of the percept? For example, the activity of some cells in a pattern may be more important than others. Does the formation of the percept depend on how features are combined? The sequential activation of multiple cells or the latency of their activation relative to brain rhythms are examples of feature combinations that may be perceptually meaningful (*1*–*3*).

The difficulty in addressing these questions is twofold. First, multiple features covary with perceptual responses, making it difficult to disentangle their independent contributions to perception. Previous studies have mainly focused on correlating neural activity with perception, in which the contributions of individual features are entangled. Although diverse stimulus sets may potentially disentangle feature contributions, biophysical constraints often impose strong correlations between features, which makes finding such stimulus sets difficult if not impossible. Second, we lack a single framework for quantitatively comparing the perceptual contributions of individual and combined features. For example, we know that single neurons (*4*) or small timing differences (*5*–*7*) can affect perception, but not their relative importance and how they come together in larger patterns to produce perception.

We investigated the link between neural activity and perception using olfaction as our model system. Odor perception is correlated with complex spatiotemporal activity patterns, and these patterns can be optogenetically manipulated in mice while measuring their perceptual responses. Odor-induced activity recruits glomeruli and spatiotemporal combinations of glomerular activity correspond to specific odorants (*8*, *9*). Both the spatial identity and temporal latencies of glomeruli activated in the pattern may be perceptually meaningful (*6*, *7*) but the relative importance of each is unknown. It is also unclear what forms a perceptually meaningful combination of glomerular activation: ordered sequences aligned relative to each other (*10*–*12*), sequences aligned to sniff rhythm (*6*, *7*, *13*, *14*), or the earliest activated subset of glomeruli within a sniff rhythm (*15*, *16*).

Here, we optogenetically manipulated individual activity features independently of other features, manipulated combinations of features, and compared the effects of all manipulations under a common metric.

Results

We first characterized the basic behavioral and neural responses to optogenetic stimulation. We chronically implanted OMP-ChR2-YFP mice ($n = 6$) (*6*) with cranial windows to expose the dorsal olfactory bulb and performed optogenetic stimulation (Fig. 1A). Similar to previous reports (*11*, *17*) and consistent with known anatomy, mitral tufted (MT) cells were activated by few localized spots (Fig. 1B and fig. S1, A and B). We observed instantaneous MT cell firing rates up to ~100 Hz, with excitatory responses lasting ~80 ms, comparable to odor-evoked responses (*13*). All spots at the same stimulation parameters were perceptually detectable but only for ChR2-positive mice (Fig. 1C) and without systematic spatial biases (Fig. 1D). We then used the same basic stimulation parameters for the main experiments.

We trained mice to discriminate Target versus Non-target synthetic odors on a two-choice task (Fig. 1E). Although the stimulation patterns we used are unlikely to correspond to specific known chemical odorants, our patterns follow stimulation durations and activation latencies that fall within known odor-evoked distributions of glomerular activation (*8*). Mice learned to discriminate a fixed Target from any random Non-target pattern over weeks, whereas ChR2-negative mice did not (Fig. 1F).

We then performed systematic spatial and/or temporal perturbations within Target patterns on a small number (10%) of probe trials and measured perceptual responses. Perturbations consisted of replacing spots with Non-target spots or shifting spots in time. For any set of perturbations, we measured the fraction of trials in which the mouse made a lick choice toward the water spout associated with the Target ("like-Target" response) as opposed to the Non-target spout. This measurement reflects perceptual distances (*18*, *19*): the perceived difference between the perturbed and original Target pattern. The larger the perceptual distance, the lower the fraction of like-Target responses. We performed a basic check to verify that perceptual distances increased as Target patterns were altered. In a pilot group of mice trained on the task, we systematically decreased the laser intensity below the intensity used during training. As the laser intensity for Target patterns was lowered, animals showed a graded decrease in like-Target responses (Fig. 1G).

Spatial perturbations led to graded changes in perceptual distances that were dependent on the timing of the spot perturbed. To perform spatial perturbations, we kept Target pattern timing fixed while replacing spots with random Non-target spots (Fig. 2A). One or more spots in the Target patterns were replaced in every possible combination of the six spot positions. We found a graded relationship between the number of spots replaced and perceptual distance (Fig. 2B). Spot replacement effects also depended on spot timing: Replacing earlier-activated spots had a greater effect than later-activated spots in both single-spot and multiple-spot replacement trials (Fig. 2C). We refer to this temporal

[1]Neuroscience Institute, NYU Langone Health, New York, NY 10016, USA. [2]Neural Computation Laboratory, Istituto Italiano di Tecnologia, 38068 Rovereto, Italy. [3]CIMeC, University of Trento, Rovereto, Italy. [4]Center for Neural Science, New York University, New York, NY 10003, USA. [5]Tech4Health Institute, NYU Langone Health, New York, NY 10010, USA. [6]Department of Ophthalmology, NYU Langone Health, New York, NY 10017, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: rinberg@nyu.edu

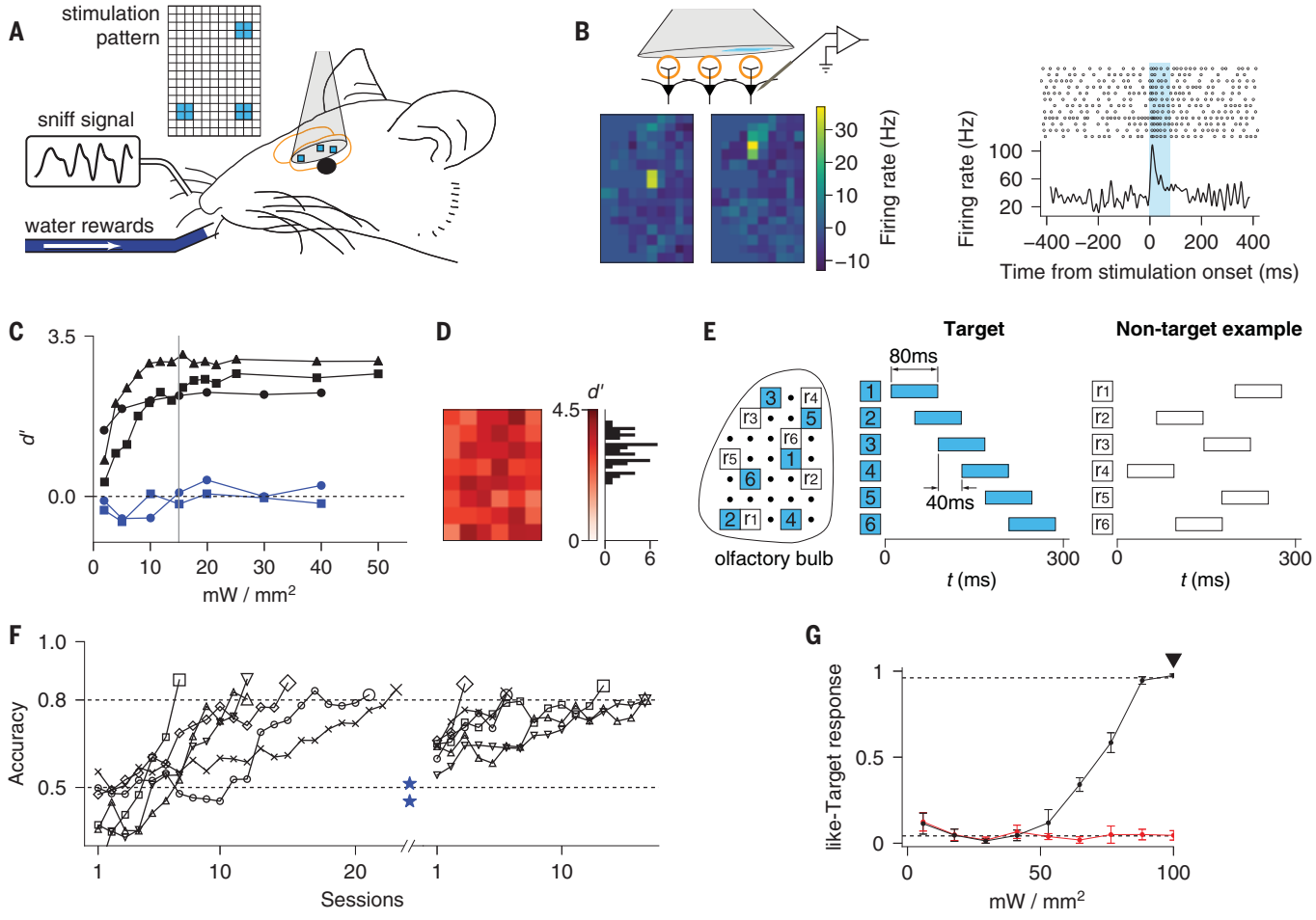


Fig. 1. Optogenetic stimulation as synthetic odors. (**A**) Schematic of the experimental setup. Dorsal olfactory bulb was exposed by a chronically implanted 3-mm window. Spatiotemporal stimulation patterns, created by a digital micromirror device, were projected onto the olfactory bulb of a head-fixed mouse in front of a pressure sensor for sniff monitoring, and lick spouts delivering water. (**B**) MT cell responses to single-spot stimulation (120 × 120 μm^2, 80-ms duration, 15 mW/mm^2) across a grid of spots. Left: Heatmaps of evoked firing (average firing rate across the stimulus duration, 80 ms) for each stimulated spot position for two typical cells. Right: Trial-by-trial spiking activity and peristimulus time histogram corresponding to left cell in (B) for the spot that evoked the largest averaged response. (**C**) Behavioral detection (d', or sensitivity index) of single-spot stimulation at different laser intensities for ChR2-positive animals (black) or ChR2-negative animals (blue). Gray vertical line marks 15 mW/mm^2. (**D**) Left: Heatmap of d' for spot detection at 15 mW/mm^2 stimulation for one ChR2 mouse at different spot positions. Right: Histogram of d' across spots. (**E**) Schematics for pattern discrimination task. Animals were trained to recognize Target versus Non-target patterns defined on a stimulation grid. Target patterns comprised six spots, initialized randomly but fixed across subsequent sessions, activated in an ordered sequence defined in time where 0 marks inhalation onset. Non-target patterns were six off-Target spots, randomly chosen from trial to trial, with randomized timing within 300 ms from inhalation (~single sniff). (**F**) Learning curves on pattern discrimination task. Mice were initially shaped by discriminating one Target pattern versus one Non-target pattern (left) to criterion performance of 0.8, then trained on the same Target versus multiple Non-target patterns (right). Note that performance below chance level (0.5) for some mice at the beginning of training is explained by a strong initial side bias (see the materials and methods). Blue stars indicate performance of control, ChR2-negative mice after >40 sessions of training. (**G**) Effect of reducing laser intensity on pattern recognition for mice previously trained to discriminate Target versus Non-target patterns at 100 mW/mm^2 (arrowhead). Shown is the probability of like-Target responses (mean ± SEM across mice) on probe trials with randomly chosen low-intensity Target patterns (black) and Non-target patterns (red). Horizontal dashed lines correspond to baseline responses to Target and Non-target patterns.

dependence as "primacy" (*15*). We additionally found that the effect of spot replacement depends on spatial distance on the olfactory bulb (fig. S2A). For these measurements, we used a larger dataset including other animals (fig. S3, A, C, and D). On average, the perceptual distance was larger between a Target and a Target with one spot replaced by a more distant spot. Because we did not measure spatial perturbations of Target locations at systematically defined locations across the olfactory bulb, our results reveal some chemotopy, but we cannot make conclusions about its characteristic length scale. Our effect was averaged across Target spot locations but could be location dependent. For example, if a Target spot was situated near to (far from) the boundary between odor receptor domains (*20*), replacing the spot with nearby spots would have large (small) effects. Our results are consistent with studies that suggest broad odor domains (*21*, *22*) and others that rule out fine-scale chemotopy (*23*–*25*). We also cannot completely rule out the alternative possibility that these effects may arise from activating fibers of passage, with strength of activation dependent on distance. However, these effects are likely to be weak: Mitral cell responses revealed only weak and spatially confined effects of fibers of passage (Fig. 1B and fig. S1), similar to other results using comparable stimulation parameters

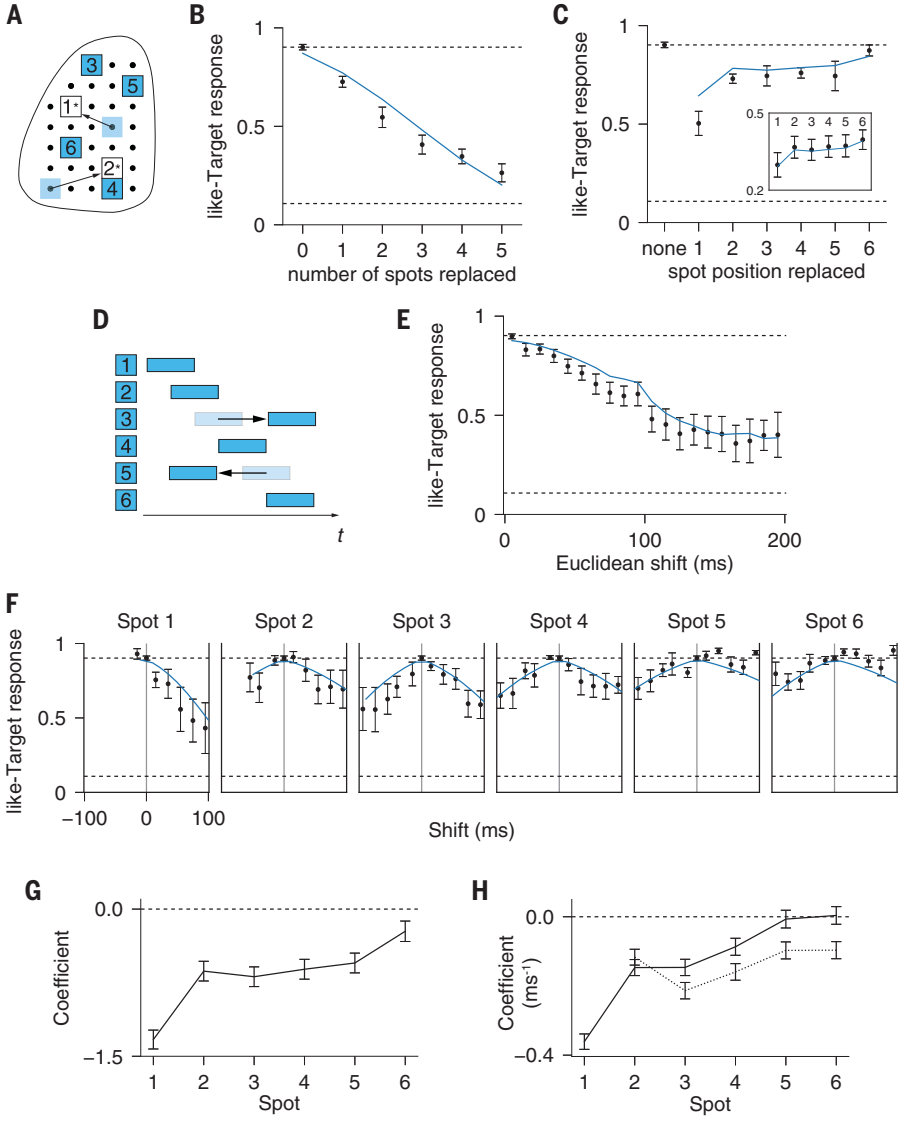


Fig. 2. Perceptual effects of spatial and temporal perturbations. (**A**) Illustration of spatial perturbations: One or multiple spots in Target patterns were randomly replaced with Non-target spots. (**B**) Probability of like-Target responses as a function of number of replaced spots in Target patterns averaged across animals. Here and later in plots of like-Target responses, horizontal dashed lines correspond to baseline responses to Target and Non-target patterns and blue lines show the corresponding regression fits. (**C**) Probability of like-Target responses as a function of a spot position for single-spot replacement. Inset: Probability of like-Target responses as a function of a spot position marginalized over trials with multiple spots replaced. (**D**) Illustration of temporal perturbations in which one or multiple spots in Target patterns were temporally shifted. (**E**) Probability of like-Target responses as a function of the amount of temporal shift ("Euclidean shift") in the pattern. (**F**) Effect of single-spot shifts plotted by spot position. (**G**) Regression coefficients showing the effect of replacing each spot in all trials, including single or multiple spot replacement. Negative coefficients imply that replacing the spot lowers like-Target responses; coefficients of zero imply no effect on responses. Regression coefficients represent the change in log-odds per unit change in the corresponding variable. (**H**) Same as (G) but for temporal perturbations. Separate coefficients were fitted for positive shifts (toward the end of the sniff cycle, solid line) and negative shifts (toward the beginning of the sniff cycle, dashed line).

(*11*). Furthermore, spatial replacement effects did not show any anisotropy that reflects the anatomy of passing fibers (fig. S2B).

Temporal perturbations also led to graded changes in perceptual distances dependent on the timing of the spot perturbed. To perform temporal perturbations, we shifted one or more spots of the Target pattern in 10-ms increments in randomly chosen combinations of spot and shift magnitude (Fig. 2D). Increasing the overall amount of temporal shift (defined as Euclidean shift = $\sqrt{\sum_i \Delta t_i^2}$ where Δt_i are individual spot shifts) led to an increase in perceptual distance (Fig. 2E). Similar to spatial perturbations, temporal perturbations also exhibited primacy. In trials in which only one spot was shifted, shifting earlier-activated spots had stronger effects than shifting later-activated spots (Fig. 2F). Furthermore, we observed an asymmetry in temporal shift effects: Shifting spots earlier, toward the beginning of the sniff cycle, had a stronger effect than shifting spots later, toward the end of the sniff cycle. This asymmetry is consistent with primacy: Although later-activated spots were less important perceptually, shifting them earlier may interfere with the early, more important spots, hence inducing larger behavioral effects.

To quantify the effect of spatial and temporal perturbations, we performed regression analyses using either binary variables for spot replacement (x_i = 1, if the spot is replaced and x_i = 0, if not) or continuous variables for temporal shifts (Δt_i). The regression models quantified primacy in spatial perturbations (Fig. 2G) and primacy and asymmetry for temporal perturbations (Fig. 2H). Primacy was reflected in larger-magnitude coefficients for earlier spots; temporal asymmetry was reflected in larger-magnitude coefficients for shifting spots earlier as opposed to later. Responses of individual mice are plotted in fig. S4.

The effects of temporal perturbations suggest that animals encode spot latencies referenced to sniff: As spots were shifted with respect to inhalation onset, animals' responses changed. Another possibility is that spot latencies are encoded with respect to other spots in the pattern. To test this, we considered the synchronous shift of the Target pattern, where all spots are shifted by the same amount with respect to inhalation (Fig. 3A). Synchronous shifts alter sniff-referenced timing while keeping pattern-referenced timing constant. Perceptual distances varied, but only weakly, with synchronous shifts (Fig. 3B). We attempted to predict the effect of any synchronous shift by summing the effects of single spots shifted by equal amounts referenced to inhalation. The prediction markedly overestimated the weak perceptual effect of synchronous shifts (Fig. 3B), suggesting an important role for pattern-referenced timing. Pattern-referenced timing can also be reflected in the order or sequence of spot activation. Order-changing temporal perturbations had a larger effect than order-preserving perturbations, controlling for equivalent amounts of sniff-referenced shift from the Target (Fig. 3C).

To explain the whole spectrum of behavioral results, we propose a spatiotemporal template-matching (STM) model. Template matching is the general process of comparing new inputs to

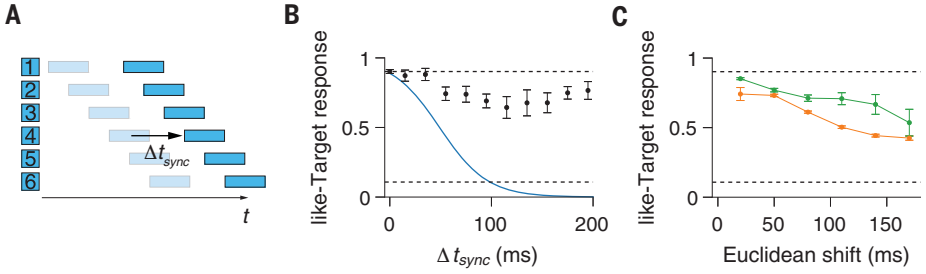


Fig. 3. Animals encode pattern-referenced timing. (**A**) Illustration of temporal perturbations in which Target patterns were shifted synchronously from inhalation by Δt_{sync}. (**B**) Like-Target response as a function of Δt_{sync} (black), with the predicted response from summing up effects of single spots shifted relative to inhalation (blue). (**C**) Comparison of average responses to temporal perturbations that preserve (green) and do not preserve (orange) temporal spot sequence for the same amount of Euclidean shift.

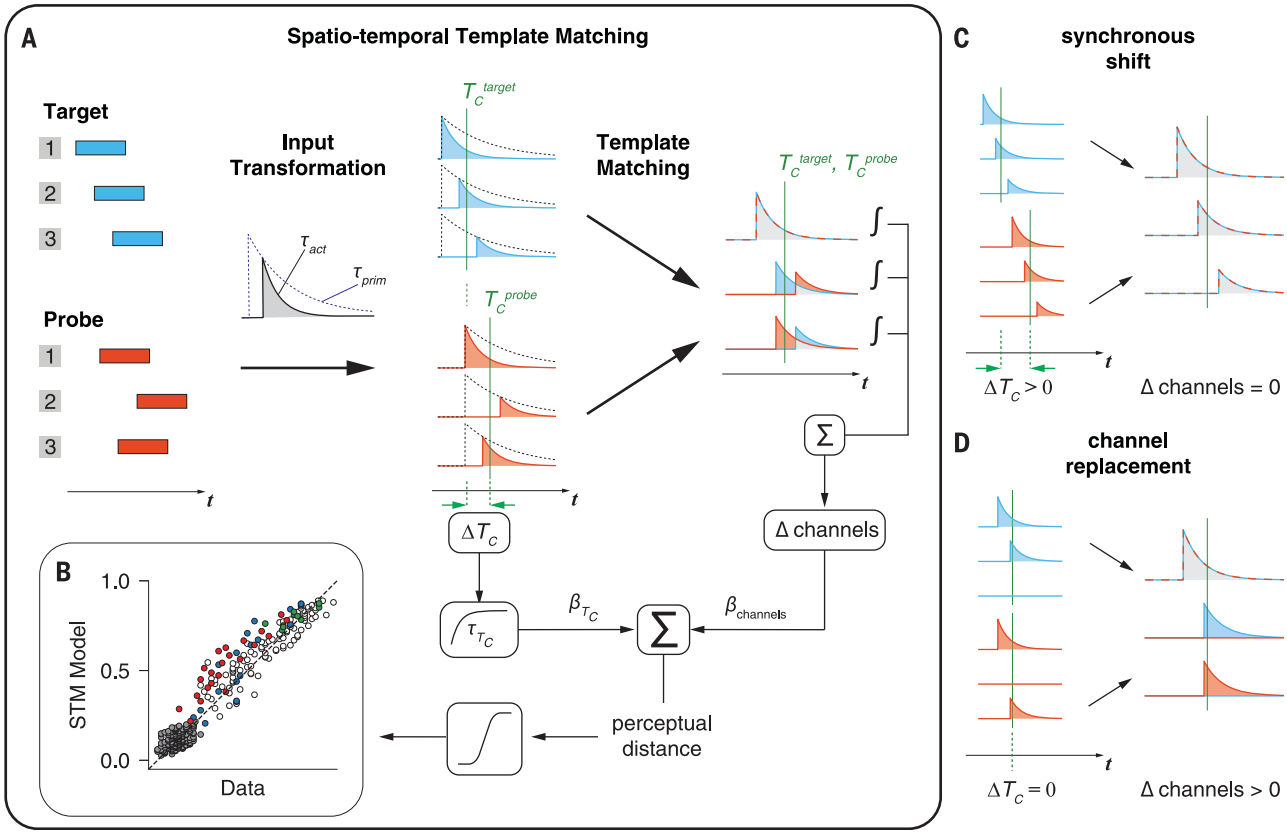


Fig. 4. STM maps olfactory inputs to perceptual outputs. (**A**) Learned pattern stored in memory (Target, blue) compared with a new input pattern (probe, red). Patterns are transformed into waveform representations: Pulse activations convolved with exponentials with decay of activation parameter τ_{act} and an amplitude that is a different exponential function (τ_{prim}) of the pulse timing from the onset of the pattern. For each set of waveforms, the center of activity, T_C^{target} and T_C^{probe}, and their absolute difference, $\Delta T_C = |T_C^{target} - T_C^{probe}|$, are calculated. Individual waveforms are then aligned to their respective pattern's center of activity and the difference in nonoverlapping area per waveform channel is integrated. This nonoverlapping area is summed linearly across all channels (Δchannels). ΔT_C is exponentiated, weighted, and then added to the weighted Δchannels value, producing a perceptual distance value. A logistic function on perceptual distances produces a probability of making a like-Target response to the given probe pattern. (**B**) STM model predictions against mouse responses across all trial types. Within each trial type, trials were sorted by perceptual distance in the model and grouped. Each dot represents a set of 50 trials. The dashed unity line indicates perfect prediction. Different colors indicate different types of perturbation: spatial, blue; synchronous temporal shifts, green; all other temporal shifts, white; Non-targets, gray; and spatial and temporal perturbation in same pattern, red. The model has the following fitting parameters: τ_{act}, τ_{prim}, τ_{T_C}, $\beta_{channels}$, β_{T_C}, and bias term β_0. (**C**) Example of template matching for synchronously shifted patterns. T_C changes, but there are no waveform differences. (**D**) Template matching for patterns in which activation in one channel is replaced by another channel. T_C does not change but waveform differences arise.

templates stored in memory (*26*). In systems neuroscience, template matching can refer to methods used by experimenters to compare spatiotemporal neural population activity (*27*–*29*). We propose that olfactory circuits compute similar forms of matching of neural activity for odor recognition.

The STM model consists of input transformation followed by template matching (Fig. 4A). Any olfactory input is transformed to temporal activations in discrete spatial channels (spots or, more generally, glomeruli). Activations decay over time and are represented by exponential waveforms. The decay slope reflects animals' sensitivity to temporal perturbations, with slower decays corresponding to higher invariance to temporal shifts. The strength of activations is further multiplicatively modulated by primacy, such that earlier-activated channels within the pattern evoke larger waveforms. Each learned Target input is transformed and stored as a template against which new patterns are compared.

A template-matching procedure then computes the perceptual distance between new patterns and Target templates. Template matching considers two effects independently: the difference in the center of activity and differences within individual spatial channels. The center of activity captures the average position of each template within a sniff cycle; similar centroid measures have been defined to characterize the mean response latency across multiple neurons (*30*). To perform within-channel comparisons, the model measures the amount of nonoverlap between waveforms. This comparison is done in reference to each template's center of activity. Consider synchronous shifts of Target patterns (Fig. 3B): Only the center of activity changes, not waveforms relative to the center of activity (Fig. 4C). Therefore, within-channel comparisons are primarily sensitive to changes in pattern-referenced timing as opposed to sniff-referenced timing. The choice of using a centroid model for pattern-referenced timing was further corroborated by regression analyses showing that the centroid model best predicts responses to temporal perturbations compared with other models (fig. S5). Within-channel comparisons are also sensitive to differences in spatial activations, for example, in the spot replacement of patterns (Fig. 4D). The nonoverlap per channel is then linearly summed across all channels to produce an overall channel difference value. Channel differences and the difference in center of activity are separately weighted and then combined to produce perceptual distances.

The STM model succeeded at predicting animals' responses to the whole spectrum of pattern perturbations (Fig. 4B) and reproduced the observed trends for individual perturbations (Fig. 5A). The model captured the graded effects of both spatial and temporal perturbations, with primacy in spatial perturbations

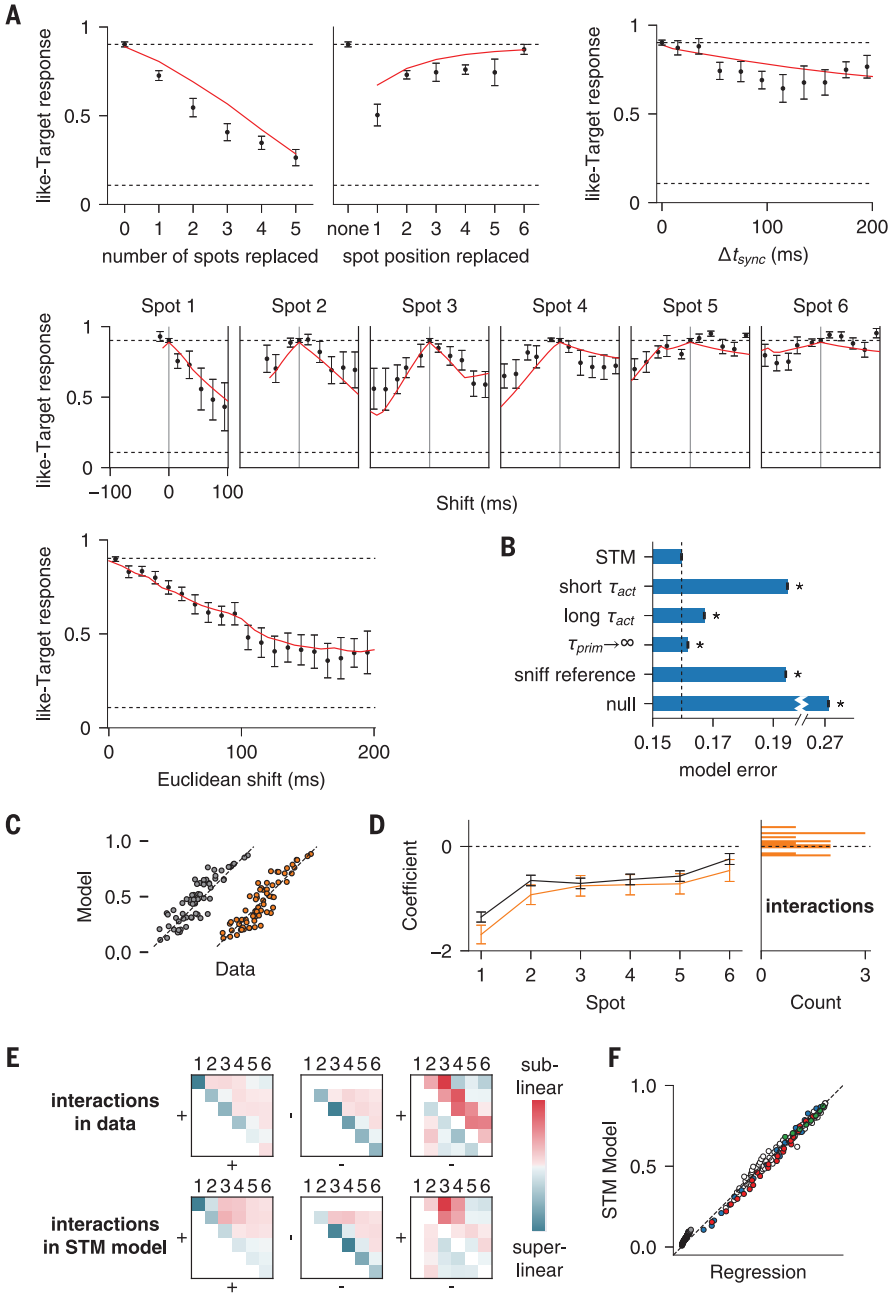


Fig. 5. Validation of the STM model. (**A**) Fits of the STM model (red) to individual spatial or temporal perturbations. (**B**) STM model compared against modified models. Vertical dashed line indicates prediction error of STM model. Short τ_{act} and long τ_{act}, changing τ_{act} from the fitted value (~60 ms) to 10 or 200 ms, respectively. $\tau_{prim} \rightarrow \infty$, setting τ_{prim} to a large value, essentially removing the primacy effect. sniff reference, within-channel comparisons referenced to sniff and not center of activity. null, null model. Asterisks indicate a $p < 0.001$ significant difference in model prediction error versus the STM model obtained from analysis of variance corrected for multiple comparisons. (**C**) Comparison of regression models for spatial perturbations, with nonlinear interactions (orange) or without (gray). The unity line (dashed) indicates perfectly predicted responses. Each dot represents a specific combination of spatial replacement by spot position. (**D**) Coefficients for models in (C) for model with interactions (orange) or without (black) plotted by spot position replaced. Right: histogram of coefficient values for interaction terms. (**E**) Comparison of temporal interactions in data as quantified by regression (top) and in simulated data run through the STM model (bottom). We considered pairwise interactions of directional shifts of spots (numbered). Left: shifting two spots later (+,+); middle: shifting two spots earlier (−,−); and right: shifting one spot later (+) and another spot earlier (−). Colors indicate magnitude and direction of interaction; red and teal indicate sublinear and superlinear interactions, meaning that the effect of paired shift is less or more than predicted from the linear sum, respectively. (**F**) Predictions of the STM model plotted against regression models for all data for the same grouping of trials as in Fig. 4B.

and primacy and asymmetry in temporal perturbations. The model also captured the weak effects of synchronous shifts. Changing any component in the model decreased model performance (Fig. 5B).

The STM model essentially performs a linear readout of spatial channels, because within-channel differences are summed up linearly without added interactions between channels. If animals respond to patterns according to model predictions, then the observed effects of spatial perturbations should also sum linearly. This is because replacing spots while keeping timing fixed produces only changes within channels (which are summed linearly) (Fig. 4D). We therefore compared a regression model containing only linear terms versus a regression model allowing for nonlinear effects. The nonlinear model did not perform better than the linear one (Fig. 5C, nonlinear/linear model errors = 0.160/0.158), and coefficients for nonlinear interactions were negligible (Fig. 5D).

Furthermore, a more general form of the STM model with an added parameter that allowed for possible interactions between spatial channels did not explain the data any better (see the supplementary text). Therefore, to the first approximation, the olfactory system performs a linear readout of spatial glomerular activity: Glomeruli contribute additively to the overall percept largely independently of the identity of the other activated glomeruli in the pattern.

On the other hand, the STM model predicts strong temporal interactions between channel activations: Shifting any spot changes the latencies of all other spots encoded in reference to the whole pattern. Therefore, the effect of shifting any spot depends on how other spots are shifted. We confirmed the existence of temporal interactions with regression: A regression model allowing temporal interactions between spots (Eq. 4) was superior to one that did not (model error = 0.1647 versus 0.1725, $p < 0.001$, t test). The STM model run on a large simulated dataset of random temporal perturbations produced the same pattern of temporal interactions in the data, as revealed by the regression model (Fig. 5E, $r = 0.83$).

We quantitatively compared the STM model with the best regression model fitted separately for each individual class of spatial or temporal perturbations. The regression models form a standard for comparison because they place minimal prior assumptions on the structure of the data. Despite the STM model having fewer parameters than the regressions (six versus 21), the STM model performed slightly better (model error = 0.159 versus 0.160 $p < 0.001$, t test). Furthermore, for any given pattern, the STM model and regression models produced responses that were tightly correlated (Fig. 5F), and the correlation between models was greater than the correlation between the STM model and the data ($r = 0.997$ versus $r = 0.97$). Therefore, the STM model produces patterns of responses similar to regression analyses, which could in principle capture more complex relationships within the data.

We further validated the STM model with unseen classes of spatiotemporal perturbations. In our described experiments and model fitting so far, Target pattern perturbations were either spatial or temporal, not both. We tested the fitted STM model on additional patterns containing both temporal shifts and spatial replacements within the same pattern. The model could predict animals' responses to spatiotemporal perturbations despite never encountering these manipulations during model fitting (Fig. 4B, prediction error on spatiotemporal perturbations = 0.23, spatial perturbations alone = 0.23, temporal perturbations alone = 0.21). Another form of spatiotemporal perturbations was inherent in Non-target trials, in which all Target spots were replaced and randomly shifted. No Non-target pattern was ever repeated for any animal, guaranteeing that all Non-target stimuli in model fitting were distinct from Non-target stimuli in model testing. The STM model predicted responses across different Non-targets (Fig. 4B, error = 0.10).

Our results did not merely reflect reinforcement learning of rewarded versus unrewarded stimulus (Target/Non-target) statistics, because a classifier trained to use all available reward information in the stimulation patterns did not reproduce animals' responses (fig. S6). The results also cannot be explained by animals responding idiosyncratically to probe trials (from lack of reward) because fitting the STM model to individual mice or to other groups of mice with different stimulus parameters (fig. S3) yielded similar results.

STM of olfactory bulb activity, linearly summed, constrained by activity-centered timing, and with preferential weighting on early activations, can account for how the olfactory system maps complex activity to perceptual outputs.

Discussion

We developed a framework for finding the perceptually meaningful subspace of spatiotemporal sensory activity. We optogenetically manipulated features in glomerular activity independently of other features, with manipulations that finely and extensively tiled spatiotemporal dimensions in a model-agnostic manner. By comparing all manipulations under a common metric, we derived a unifying, model-based account of olfactory computations. Other studies have used similar optogenetic techniques to probe olfaction, manipulating single-feature dimensions while measuring downstream neural responses (*11*, *17*, *31*) or the limits of behavioral discrimination (*6*, *7*, *12*). Those studies established which single-activity features are potentially accessible to the olfactory system, whereas our approach allows us to understand how features are organized in complex patterns to generate perception.

We measured how perceptual responses vary across systematic input manipulations, a basic approach in psychology applied to various classes of stimuli (e.g., images, sounds) to reveal underlying models of sensory representation (*18*, *19*). We derived models of neural representation by measuring perceptual responses across changes in neural activity.

The STM model resolves several open issues in olfactory coding. First, the model suggests a linear readout of spatial glomerular patterns. It has been proposed that odors are encoded in specific spatial combinations of glomeruli (*32*), but there are multiple possible combinatorial coding schemes that have not been tested. For example, one extreme possibility is a "barcode" representation in which any slight change to the combination leads to a completely unrelated odor, maximizing the system's representational capacity for odors. Instead, we found a perceptual readout of spatial patterns that is linear. This is consistent with studies that imaged odor-evoked glomerular activity (*33*–*35*) or directly stimulated individual glomeruli (*36*). A linear readout implies that two patterns will generate odors that are perceptually similar depending on the degree of glomerular overlap. Although a linear readout has lower representational capacity, it may explain the generalization of odor percepts across varying concentrations or backgrounds.

The role of temporal sequences in odor perception has long been hypothesized (*8*, *13*) but is not theoretically required to support odor perception (*37*). We provide direct experimental evidence for the role of temporal sequences: Mice trained to recognize a synthetic odor pattern use temporal sequences in odor recognition even though other (spatial) cues are sufficient to solve the task. These sequences were defined with relative latencies within a pattern (*10*), as opposed to latencies with respect to sniff (*6*, *8*, *13*). Pattern-referenced timing may reflect how inputs compete or are integrated downstream because competition or integration can depend on temporal proximity between inputs (*11*, *38*, *39*). Biologically plausible mechanisms for reading out these pattern-referenced sequences may involve delay circuits (*40*–*42*). Long delays on the order of 100 ms can be implemented by recurrent circuitry: In piriform cortex, recurrent excitation (*43*, *44*) or inhibition (*43*, *45*–*47*) can potentially generate sensitivity to long delays in olfactory inputs (*11*, *48*).

We found a weak perceptual effect for changing the overall position (center of activity) of the pattern within sniff, possibly arising from weak modulation of glomerular activity by sniff-coupled mechanosensory responses (*14*).

Alternatively, mice may be using the overall position of the pattern in sniff as a weak cue to solve the task, because the average randomly generated Non-target pattern has a different position within sniff compared with the Target pattern.

We found a primacy effect in which earlier-activated glomeruli had larger effects on perceptual responses. Primacy has been suggested as a strategy for animals to recognize the same odor across varying concentrations, because early-activated glomeruli remain stable across different concentrations. The ability of animals trained to recognize odors across varying concentrations was impaired during coarse optogenetic disruption, but only when the disruption occurred in the initial ~100 ms of inhalation (*15*). In the present study, animals were trained to recognize a single Target synthetic odor and displayed a primacy effect without explicit training to "concentration variants" of the Target. Therefore, the preferential weighting of early inputs occurs under different task demands, suggesting that primacy is a fundamental property of the olfactory system. Primacy may be supported by downstream computations at the level of mitral cells (*49*) or piriform cortex (*16*, *50*).

In our task, the animal could use multiple neural features to form a decision boundary for Target/Non-target discrimination because we did not tune task difficulty to the limit of perception where decision boundaries can be more confined. Therefore, we cannot rule out the possibility that nonsensory processes such as decision-making or action-related processes contributed to the observed responses, and the STM model might misattribute some of these more downstream phenomena to sensory processing alone.

We cannot rule out the possibility that fibers of passage were activated during our spot stimulation. However, this effect is likely to be weak; MT cells were activated by few localized spots at the laser intensity chosen for our experiments. Another study using similar stimulation parameters also found negligible effects of fibers of passage (*11*). Furthermore, we observed negligible nonlinear interactions between spots during spatial replacements, which would not be expected if there were widespread coactivation of glomeruli from fibers of passage.

Our experiments did not explore the dependence of perceptual similarity on the chemical tuning of glomeruli. Various studies have suggested several chemical dimensions that may correlate with perception (*51*, *52*). If perceptual similarity between glomeruli depends on chemical similarity along specific dimensions, then chemical similarity may be an additional variable that modulates interactions between glomeruli in the STM model.

Our synthetic stimuli may not capture the full complexity of glomerular activity evoked by natural odors. We view the synthetic approach as complementary to approaches using more naturalistic stimuli. The claim is not that these synthetic stimuli are direct proxies for naturalistic stimuli, but rather that synthetic stimuli afford well-controlled experiments with precise parameterization and causal manipulation, and can be used to establish basic principles of the neural code. In other well-studied sensory systems, foundational understanding of sensory processing has been built upon synthetic, reduced stimuli (*53*), whereas naturalistic stimuli have been used to test and refine foundational models (*54*).

We developed an experimental and theoretical approach that links the complex spatiotemporal language of the brain with perception and behavior. This is especially pertinent given continued advancements in optical techniques that allow direct access and manipulation of the spatiotemporal neural codes at the level of computations consequential to behavior (*55*–*58*).

Methods summary

We used OMP-ChR2 mice in which ChR2 is expressed in all olfactory glomeruli in the olfactory bulb. Dorsal olfactory bulb was exposed for optogenetic stimulation with chronically implanted glass coverslips. In separate neural recording experiments, mice were implanted with glass windows containing pre-drilled, ~1-mm holes, which allowed insertion of silicon probes for measuring mitral cell responses. For patterned optogenetic stimulation, we used digital micromirror devices in optical setups that were either purchased commercially or custom-built, allowing stimulation patterns with up to 10-μm spatial and 1-ms temporal resolution. Stimulus presentation and the recording of behavioral variables such as respiration and lick responses were performed with a custom behavioral control system based on Arduino microcontrollers. For the main behavioral experiment, mice were trained on a two-alternative forced choice task in which one lick spout was assigned to a Target pattern and the other was assigned to Non-target patterns, and mice were rewarded with a water droplet for licking to the correct spout when the associated pattern was presented. A partial reinforcement schedule was used in which only 70% of all trials were rewarded. A subset of these unrewarded trials were later substituted with probe trials (perturbations of Target patterns), such that probe trials were never rewarded regardless of the animal's response. Animals' responses on each trial were then recorded as binary responses: licking to the Target-associated spout or Non-target–associated spout. We fitted this binary response data using logistic regression models to quantify the behavioral effect along each spatial or temporal perturbation dimension. The STM model was then used to account for responses to all perturbation dimensions, both individually and combined.

REFERENCES AND NOTES

1. M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge Univ. Press, 1991).
2. G. Buzsáki, Neural syntax: Cell assemblies, synapsembles, and readers. *Neuron* **68**, 362–385 (2010). doi: 10.1016/j.neuron.2010.09.023; pmid: 21040841
3. C. A. Runyan, E. Piasini, S. Panzeri, C. D. Harvey, Distinct timescales of population coding across cortex. *Nature* **548**, 92–96 (2017). doi: 10.1038/nature23020; pmid: 28723889
4. A. R. Houweling, M. Brecht, Behavioural report of single neuron stimulation in somatosensory cortex. *Nature* **451**, 65–68 (2008). doi: 10.1038/nature06447; pmid: 18094684
5. Y. Yang, M. R. DeWeese, G. H. Otazu, A. M. Zador, Millisecond-scale differences in neural activity in auditory cortex can drive decisions. *Nat. Neurosci.* **11**, 1262–1263 (2008). doi: 10.1038/nn.2211; pmid: 18849984
6. M. Smear, R. Shusterman, R. O'Connor, T. Bozza, D. Rinberg, Perception of sniff phase in mouse olfaction. *Nature* **479**, 397–400 (2011). doi: 10.1038/nature10521; pmid: 21993623
7. M. Smear, A. Resulaj, J. Zhang, T. Bozza, D. Rinberg, Multiple perceptible signals from a single olfactory glomerulus. *Nat. Neurosci.* **16**, 1687–1691 (2013). doi: 10.1038/nn.3519; pmid: 24056698
8. R. M. Carey, J. V. Verhagen, D. W. Wesson, N. Pírez, M. Wachowiak, Temporal structure of receptor neuron input to the olfactory bulb imaged in behaving rats. *J. Neurophysiol.* **101**, 1073–1088 (2009). doi: 10.1152/jn.90902.2008; pmid: 19091924
9. H. Spors, A. Grinvald, Spatiotemporal dynamics of odor representations in the mammalian olfactory bulb. *Neuron* **34**, 301–315 (2002). doi: 10.1016/S0896-6273(02)00644-X; pmid: 11970871
10. S. Junek, E. Kludt, F. Wolf, D. Schild, Olfactory coding with patterns of response latencies. *Neuron* **67**, 872–884 (2010). doi: 10.1016/j.neuron.2010.08.005; pmid: 20826317
11. R. Haddad *et al.*, Olfactory cortical neurons read out a relative time code in the olfactory bulb. *Nat. Neurosci.* **16**, 949–957 (2013). doi: 10.1038/nn.3407; pmid: 23685720
12. M. R. Rebello *et al.*, Perception of odors linked to precise timing in the olfactory system. *PLOS Biol.* **12**, e1002021 (2014). doi: 10.1371/journal.pbio.1002021; pmid: 25514030
13. R. Shusterman, M. C. Smear, A. A. Koulakov, D. Rinberg, Precise olfactory responses tile the sniff cycle. *Nat. Neurosci.* **14**, 1039–1044 (2011). doi: 10.1038/nn.2877; pmid: 21765422
14. R. Iwata, H. Kiyonari, T. Imai, Mechanosensory-based phase coding of odor identity in the olfactory bulb. *Neuron* **96**, 1139–1152.e7 (2017). doi: 10.1016/j.neuron.2017.11.008; pmid: 29216451
15. C. D. Wilson, G. O. Serrano, A. A. Koulakov, D. Rinberg, A primacy code for odor identity. *Nat. Commun.* **8**, 1477 (2017). doi: 10.1038/s41467-017-01432-4; pmid: 29133907
16. K. A. Bolding, K. M. Franks, Recurrent cortical circuits implement concentration-invariant odor coding. *Science* **361**, eaat6904 (2018). doi: 10.1126/science.aat6904; pmid: 30213885
17. A. K. Dhawale, A. Hagiwara, U. S. Bhalla, V. N. Murthy, D. F. Albeanu, Non-redundant odor coding by sister mitral cells revealed by light addressable glomeruli in the mouse. *Nat. Neurosci.* **13**, 1404–1412 (2010). doi: 10.1038/nn.2673; pmid: 20953197
18. R. N. Shepard, Toward a universal law of generalization for psychological science. *Science* **237**, 1317–1323 (1987). doi: 10.1126/science.3629243; pmid: 3629243
19. A. Tversky, Features of similarity. *Psychol. Rev.* **84**, 327–352 (1977). doi: 10.1037/0033-295X.84.4.327
20. T. Bozza *et al.*, Mapping of class I and class II odorant receptors to glomerular domains by two distinct types of olfactory sensory neurons in the mouse. *Neuron* **61**, 220–233 (2009). doi: 10.1016/j.neuron.2008.11.010; pmid: 19186165
21. K. Mori, Y. K. Takahashi, K. M. Igarashi, M. Yamaguchi, Maps of odorant molecular features in the Mammalian olfactory bulb. *Physiol. Rev.* **86**, 409–433 (2006). doi: 10.1152/physrev.00021.2005; pmid: 16601265
22. B. A. Johnson, M. Leon, Chemotopic odorant coding in a mammalian olfactory system. *J. Comp. Neurol.* **503**, 1–34 (2007). doi: 10.1002/cne.21396; pmid: 17480025
23. L. Ma *et al.*, Distributed representation of chemical features and tunotopic organization of glomeruli in the mouse olfactory bulb. *Proc. Natl. Acad. Sci. USA.* **109**, 5481–5486 (2012). doi: 10.1073/pnas.1117491109; pmid: 22431605

24. E. R. Soucy, D. F. Albeanu, A. L. Fantana, V. N. Murthy, M. Meister, Precision and diversity in an odor map on the olfactory bulb. *Nat. Neurosci.* **12**, 210–220 (2009). doi: 10.1038/nn.2262; pmid: 19151709
25. J. Zhang, G. Huang, A. Dewan, P. Feinstein, T. Bozza, Uncoupling stimulus specificity and glomerular position in the mouse olfactory system. *Mol. Cell. Neurosci.* **51**, 79–88 (2012). doi: 10.1016/j.mcn.2012.08.006; pmid: 22926192
26. U. Neisser, in *Cognitive Psychology: Key Readings*, D. A. Balota, E. J. Marsh, Eds. (Psychology Press, 2004), pp. 125–150.
27. C. Houghton, K. Sen, A new multineuron spike train metric. *Neural Comput.* **20**, 1495–1511 (2008). doi: 10.1162/neco.2007.10-06-350; pmid: 18194108
28. A. K. Lee, M. A. Wilson, Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194 (2002). doi: 10.1016/S0896-6273(02)01096-6; pmid: 12495631
29. K. Louie, M. A. Wilson, Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* **29**, 145–156 (2001). doi: 10.1016/S0896-6273(01)00186-6; pmid: 11182087
30. A. Luczak, P. Barthó, K. D. Harris, Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* **62**, 413–425 (2009). doi: 10.1016/j.neuron.2009.03.014; pmid: 19447096
31. I. G. Davison, M. D. Ehlers, Neural circuit mechanisms for pattern detection and feature combination in olfactory cortex. *Neuron* **70**, 82–94 (2011). doi: 10.1016/j.neuron.2011.02.047; pmid: 21482358
32. B. Malnic, J. Hirono, T. Sato, L. B. Buck, Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999). doi: 10.1016/S0092-8674(00)80581-4; pmid: 10089886
33. S. A. Kreher, D. Mathew, J. Kim, J. R. Carlson, Translation of sensory input into behavioral output via an olfactory system. *Neuron* **59**, 110–124 (2008). doi: 10.1016/j.neuron.2008.06.010; pmid: 18614033
34. D. Rokni, V. Hemmelder, V. Kapoor, V. N. Murthy, An olfactory cocktail party: Figure-ground segregation of odorants in rodents. *Nat. Neurosci.* **17**, 1225–1232 (2014). doi: 10.1038/nn.3775; pmid: 25086608
35. A. Mathis, D. Rokni, V. Kapoor, M. Bethge, V. N. Murthy, Reading out olfactory receptors: Feedforward circuits detect odors in mixtures without demixing. *Neuron* **91**, 1110–1123 (2016). doi: 10.1016/j.neuron.2016.08.007; pmid: 27593177
36. J. S. Bell, R. I. Wilson, Behavior reveals selective summation and max pooling among olfactory processing channels. *Neuron* **91**, 425–438 (2016). doi: 10.1016/j.neuron.2016.06.011; pmid: 27373835
37. A. Koulakov, A. Gelperin, D. Rinberg, Olfactory coding with all-or-nothing glomeruli. *J. Neurophysiol.* **98**, 3134–3142 (2007). doi: 10.1152/jn.00560.2007; pmid: 17855585
38. K. M. Franks, J. S. Isaacson, Strong single-fiber sensory inputs to olfactory cortex: Implications for olfactory coding. *Neuron* **49**, 357–363 (2006). doi: 10.1016/j.neuron.2005.12.026; pmid: 16446140
39. A. Apicella, Q. Yuan, M. Scanziani, J. S. Isaacson, Pyramidal cells in piriform cortex receive convergent input from distinct olfactory bulb glomeruli. *J. Neurosci.* **30**, 14255–14260 (2010). doi: 10.1523/JNEUROSCI.2747-10.2010; pmid: 20962246
40. J. J. Hopfield, Pattern recognition computation using action potential timing for stimulus representation. *Nature* **376**, 33–36 (1995). doi: 10.1038/376033a0; pmid: 7596429
41. H. Sanders *et al*., A network that performs brute-force conversion of a temporal sequence to a spatial pattern: Relevance to odor recognition. *Front. Comput. Neurosci.* **8**, 108 (2014). doi: 10.3389/fncom.2014.00108; pmid: 25278870
42. D. W. Tank, J. J. Hopfield, Neural computation by concentrating information in time. *Proc. Natl. Acad. Sci, USA.* **84**, 1896–1900 (1987). doi: 10.1073/pnas.84.7.1896; pmid: 3470765
43. K. M. Franks *et al*., Recurrent circuitry dynamically shapes the activation of piriform cortex. *Neuron* **72**, 49–56 (2011). doi: 10.1016/j.neuron.2011.08.020; pmid: 21982368
44. C. Poo, J. S. Isaacson, A major role for intracortical circuits in the strength and tuning of odor-evoked excitation in olfactory cortex. *Neuron* **72**, 41–48 (2011). doi: 10.1016/j.neuron.2011.08.015; pmid: 21982367
45. M. Satou, K. Mori, Y. Tazawa, S. F. Takagi, Interneurons mediating fast postsynaptic inhibition in pyriform cortex of the rabbit. *J. Neurophysiol.* **50**, 89–101 (1983). doi: 10.1152/jn.1983.50.1.89; pmid: 6875654
46. C. C. A. Stokes, J. S. Isaacson, From dendrite to soma: Dynamic routing of inhibition by complementary interneuron microcircuits in olfactory cortex. *Neuron* **67**, 452–465 (2010). doi: 10.1016/j.neuron.2010.06.029; pmid: 20696382
47. N. Suzuki, J. M. Bekkers, Microcircuits mediating feedforward and feedback synaptic inhibition in the piriform cortex. *J. Neurosci.* **32**, 919–931 (2012). doi: 10.1523/JNEUROSCI.4112-11.2012; pmid: 22262890
48. L. B. Haberly, Summed potentials evoked in opossum prepyriform cortex. *J. Neurophysiol.* **36**, 775–788 (1973). doi: 10.1152/jn.1973.36.4.775; pmid: 4713319
49. E. M. Arneodo *et al*., Stimulus dependent diversity and stereotypy in the output of an olfactory functional unit. *Nat. Commun.* **9**, 1347 (2018). doi: 10.1038/s41467-018-03837-1; pmid: 29632302
50. K. A. Bolding, K. M. Franks, Complementary codes for odor identity and intensity in olfactory cortex. *eLife* **6**, e22630 (2017). doi: 10.7554/eLife.22630; pmid: 28379135
51. A. Keller *et al*., Predicting human olfactory perception from chemical features of odor molecules. *Science* **355**, 820–826 (2017). doi: 10.1126/science.aal2014; pmid: 28219971
52. K. Snitz *et al*., Predicting odor perceptual similarity from odor structure. *PLOS Comput. Biol.* **9**, e1003184 (2013). doi: 10.1371/journal.pcbi.1003184; pmid: 24068899
53. D. H. Hubel, T. N. Wiesel, Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148**, 574–591 (1959). doi: 10.1113/jphysiol.1959.sp006308; pmid: 14403679
54. N. C. Rust, J. A. Movshon, In praise of artifice. *Nat. Neurosci.* **8**, 1647–1650 (2005). doi: 10.1038/nn1606; pmid: 16306892
55. G. M. Lerman, J. V. Gill, D. Rinberg, S. Shoham, Precise optical probing of perceptual detection. bioRxiv 456764 [Preprint] 30 October 2018; https://doi.org/10.1101/456764.
56. A. M. Packer, L. E. Russell, H. W. P. Dalgleish, M. Häusser, Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. *Nat. Methods* **12**, 140–146 (2015). doi: 10.1038/nmeth.3217; pmid: 25532138
57. L. Carrillo-Reid, S. Han, W. Yang, A. Akrouh, R. Yuste, Controlling visually guided behavior by holographic recalling of cortical ensembles. *Cell* **178**, 447–457.e5 (2019). doi: 10.1016/j.cell.2019.05.045; pmid: 31257030
58. J. H. Marshel *et al*., Cortical layer-specific critical dynamics triggering perception. *Science* **365**, eaaw5202 (2019). doi: 10.1126/science.aaw5202; pmid: 31320556
59. E. Chong, M. Moroni, C. Wilson, S. Shoham, S. Panzeri, D. Rinberg, Code for: Manipulating synthetic optogenetic odors reveals the coding logic of olfactory perception, Zenodo (2020); https://doi.org/10.5281/zenodo.3698314.

ACKNOWLEDGMENTS

We thank A. Resulaj, G. Serrano, A. Pickens, S. Stark, S. Toole, H. Ma, G. Lerman, J. Kappel, N. Amin, A. Oganov, M. Robles-Long, and A. Pillai for assistance in building and running experiments and B. Mensh, K. Nagel, T. Movshon, W. Ma, J. Gill, and H. Nakayama for helpful comments. **Funding:** This work was supported by NIH BRAIN Initiative Grants R01NS109961 to D.R. and S.P. and U19NS107464 to D.R., S.P., and S.S. **Author contributions:** E.C., S.P., S.S., and D.R. conceived and designed the experimental, computational, and modeling approaches; E.C. and C.W. built the experimental system and performed experiments; M.M., C.W., and E.C. performed computational analyses; E.C., M.M., S.S., S.P., and D.R. wrote the manuscript; and S.P. and D.R. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Datasets and analysis code will be available at Zenodo (*59*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/eaba2357/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
References (*60*–*66*)

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

15 November 2019; accepted 1 May 2020
10.1126/science.aba2357

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

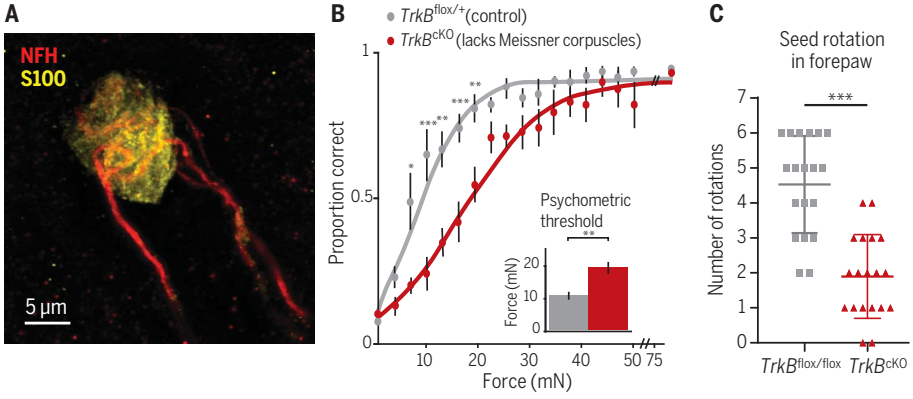
Meissner corpuscles and their spatially intermingled afferents underlie gentle touch perception

Nicole L. Neubarth*, Alan J. Emanuel*, Yin Liu*, Mark W. Springel, Annie Handler, Qiyu Zhang, Brendan P. Lehnert, Chong Guo, Lauren L. Orefice, Amira Abdelaziz, Michelle M. DeLisle, Michael Iskols, Julia Rhyins, Soo J. Kim, Stuart J. Cattel, Wade Regehr, Christopher D. Harvey, Jan Drugowitsch, David D. Ginty†

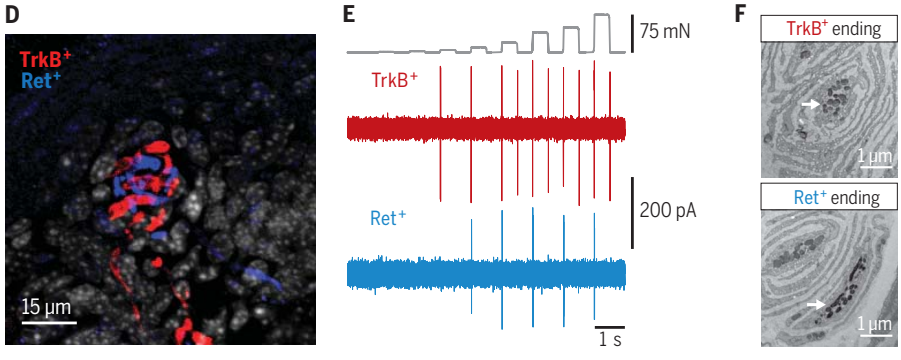
INTRODUCTION: Meissner corpuscles are mechanosensory end organs that densely occupy mammalian glabrous skin. The basic anatomy of the Meissner corpuscle and the Aβ (large soma diameter and fast action potential conducting) mechanosensory neurons that innervate it have been widely described. However, little is known about the requirement for the Meissner corpuscle and its innervating Aβ mechanosensory neurons for touch-related behaviors, sensorimotor capabilities, and tactile perception.

RATIONALE: Mice lacking either brain-derived neurotrophic factor (BDNF) or its receptor TrkB have a range of developmental deficits, including an absence of Meissner corpuscles. We reasoned that confining TrkB manipulations to sensory neurons in conjunction with selective genetic labeling approaches would allow us to investigate the developmental assembly of Meissner corpuscles and their functions in somatosensation. While examining Meissner corpuscle development, we found that two Aβ mechanosensory neuron types innervate the corpuscle. Here, we assessed the requirement of Meissner corpuscles for touch sensitivity and fine sensorimotor control and the anatomical, physiological, and ultrastructural properties of the two Aβ sensory neurons that innervate this mechanosensory end organ.

Meissner corpuscles are required for perception of gentle touch and fine sensorimotor control



Two molecularly and physiologically distinct Aβ neurons innervate Meissner corpuscles and differentially associate with corpuscle lamellar cells



Anatomy and physiology of Meissner corpuscles required for tactile behavior. (**A** to **C**) Meissner corpuscles and their Aβ afferents (A) were required for normal tactile sensitivity (B) and fine sensorimotor control (C). (**D**) TrkB$^+$ and Ret$^+$ Aβ neurons both innervated Meissner corpuscles. (**E** and **F**) TrkB$^+$ Meissner afferents were more sensitive (E) and had more lamellar cell wrappings [(F), arrows indicate axons wrapped with lamellar processes] compared with Ret$^+$ Meissner afferents. In (A), S100 and NFH are antibodies used to visualize Meissner corpuscles and their afferents, respectively. In (B) and (C), error bars represent SEM, and *p < 0.05, **p < 0.01, and ***p < 0.001.

RESULTS: Sensory neuron–specific knockout of TrkB resulted in complete loss of Meissner corpuscles without affecting other mechanosensory end organs in mouse glabrous (nonhairy) skin, including Aβ mechanosensory neuron–Merkel cell complexes and Pacinian corpuscles. Behavioral measurements showed that mice lacking Meissner corpuscles were deficient in perceiving and reacting to the gentlest detectable forces acting on glabrous skin and in fine sensorimotor control.

ON OUR WEBSITE

Read the full article at https://dx.doi.org/10.1126/science.abb2751

Genetic labeling experiments revealed that Meissner corpuscles are innervated by two molecularly distinct Aβ sensory neuron types, one that expresses TrkB and the other that expresses the tyrosine kinase Ret. Despite innervating the same end organ, the responses of these two Aβ neuron types to tactile stimuli differed: The TrkB-positive (TrkB$^+$) Meissner afferent was more sensitive and responded at both the onset and the offset of step indentations of glabrous skin, whereas the Ret$^+$ Meissner afferent was less sensitive and rarely responded at step offset. Some Ret$^+$ neurons even had sustained responses during a static indentation stimulus. In addition, the axonal endings of these two Aβ mechanosensory neuron types were found to be homotypically tiled but heterotypically offset. Computational modeling suggested that this anatomical arrangement maximizes information available for encoding acuity while ensuring complete coverage of the skin using a limited number of neurons. Finally, ultrastructural analysis using electron microscopy revealed that the axon terminals of the more sensitive TrkB$^+$ Meissner afferents had greater numbers of lamellar cell wrappings than the terminals of the less sensitive Ret$^+$ Meissner afferents.

CONCLUSION: We conclude that two homotypically tiled but heterotypically offset Aβ mechanosensory neurons with distinct molecular, physiological, and ultrastructural properties innervate Meissner corpuscles, which underlie the perception of, and behavioral responses to, the gentlest detectable forces acting on glabrous skin. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: david_ginty@hms.harvard.edu
Cite this article as N. L. Neubarth *et al.*, *Science* **368**, eabb2751 (2020). DOI: 10.1126/science.abb2751

RESEARCH ARTICLE

NEUROSCIENCE

Meissner corpuscles and their spatially intermingled afferents underlie gentle touch perception

Nicole L. Neubarth[1,2]*†, Alan J. Emanuel[1,2]*, Yin Liu[3,4]*, Mark W. Springel[1,2], Annie Handler[1,2], Qiyu Zhang[1,2], Brendan P. Lehnert[1,2], Chong Guo[1], Lauren L. Orefice[1,2], Amira Abdelaziz[1,2], Michelle M. DeLisle[1,2], Michael Iskols[1,2], Julia Rhyins[1,2], Soo J. Kim[1,2], Stuart J. Cattel[1,2], Wade Regehr[1], Christopher D. Harvey[1], Jan Drugowitsch[1], David D. Ginty[1,2]‡

Meissner corpuscles are mechanosensory end organs that densely occupy mammalian glabrous skin. We generated mice that selectively lacked Meissner corpuscles and found them to be deficient in both perceiving the gentlest detectable forces acting on glabrous skin and fine sensorimotor control. We found that Meissner corpuscles are innervated by two mechanoreceptor subtypes that exhibit distinct responses to tactile stimuli. The anatomical receptive fields of these two mechanoreceptor subtypes homotypically tile glabrous skin in a manner that is offset with respect to one another. Electron microscopic analysis of the two Meissner afferents within the corpuscle supports a model in which the extent of lamellar cell wrappings of mechanoreceptor endings determines their force sensitivity thresholds and kinetic properties.

The basic anatomy of the Meissner corpuscle and its presumed innervating Aβ rapidly adapting type I (RAI) low-threshold mechanoreceptor (LTMR) have been widely described since its discovery in 1852 (*1*). However, the Meissner corpuscle's requirement for touch-related behaviors, sensorimotor capabilities, and tactile perception have remained a matter of speculation.

Results

BDNF signaling is required for Meissner corpuscle development

Mice lacking either brain-derived neurotrophic factor (BDNF) or its receptor TrkB die neonatally and notably also have a complete absence of Meissner corpuscles (*2*–*4*). We thus generated conditional mutant mouse models that are viable in adulthood and selectively lack Meissner corpuscles for addressing their role in tactile perception and sensorimotor behaviors in adult animals. We found that BDNF is expressed in glabrous skin during the period of Meissner corpuscle formation (fig. S1, A to C). Eliminating BDNF expression in epithelial cells of the skin using *K5Cre; BDNF*$^{flox/flox}$ mice resulted in a dramatic reduction in the number of Meissner corpuscles (fig. S1, F to H). We found no disturbance of Meissner corpuscles in mice lacking BDNF in neurons of the dorsal root ganglia (DRG) or Schwann cells (*Wnt1*^{Cre}*; BDNF*$^{flox/flox}$; fig. S1, D, E, and H), suggesting that skin-derived, and not neuron- or Schwann cell–derived, BDNF is crucial for Meissner corpuscle development. BDNF's cognate receptor, TrkB, is present in a subset of primary sensory neurons of the DRG, and mice that selectively lack TrkB in primary sensory neurons (*Advillin*^{Cre}*; TrkB*$^{flox/flox}$, hereafter called *TrkB*^{cKO}) (*5*, *6*), although viable and overtly normal, are also devoid of Meissner corpuscles and their innervating sensory neurons (Fig. 1, A and B, and fig. S2). Mice that lack TrkB in Schwann cells, on the other hand, have a normal complement of Meissner corpuscles (fig. S2, H and J). Although some neurofilament-positive (NFH$^+$) sensory fibers were present in glabrous skin dermal papillae of *TrkB*^{cKO} mice before postnatal day 20 (P20), these fibers were absent in adults (Fig. 1A and fig. S2). By contrast, Merkel cells and the cutaneous endings of their associated Aβ slowly adapting (SA) LTMRs, which are TrkB-negative, were found in comparable numbers in glabrous skin of control and *TrkB*^{cKO} adult mice (Fig. 1, A and B). We next did in vivo loose-patch electrophysiological recordings (*7*, *8*) of random L4 DRG neurons from adult control and *TrkB*^{cKO} mice. Whereas indentation of hindpaw glabrous skin with gentle force steps evoked both Aβ RA-LTMR and Aβ SA-LTMR physiological responses in control mice, Aβ RA-LTMR responses were absent in *TrkB*^{cKO} mice even at indentation forces as high as 75 mN (Fig. 1, D to F). Activation thresholds and firing patterns of Aβ SA-LTMRs in control and *TrkB*^{cKO} mice were indistinguishable (Fig. 1G). Thus, BDNF and TrkB expressed in glabrous skin epithelial cells and sensory neurons of the DRG, respectively, are essential for development of Meissner corpuscles and the presence of Aβ RA-LTMR responses to indentation of glabrous skin. By contrast, sensory-neuron TrkB signaling is dispensable for the development and normal response properties of Merkel cells and their associated Aβ SA-LTMRs and, as found previously, the development of Pacinian corpuscles (*2*–*4*, *9*).

Meissner corpuscles are required for normal glabrous skin sensitivity

We subjected adult control and *TrkB*^{cKO} mice to behavioral tests of glabrous skin sensitivity, light touch perception, gait analysis, and fine sensorimotor control. First, we measured glabrous hindpaw withdrawal thresholds in *TrkB*^{cKO} and littermate control mice using forces ranging from 0.008 to 4.0 g applied to the pedal pads of the hindpaw (Fig. 2A). Wild-type mice withdrew their paws in response to the entire range of forces presented, rarely reacting to the lowest forces and nearly always reacting to the highest forces. On the contrary, although *TrkB*^{cKO} mice displayed normal hindpaw withdrawal at the high end of this force range, they were unresponsive to the lightest forces (Fig. 2A). We infer from these behavioral measurements that activation of Merkel cell complexes and Aβ SA-LTMRs does not underlie sensorimotor responses to the lightest detectable forces acting on glabrous skin. To further test whether this deficit was specific to disruption of Meissner corpuscles and not to disturbance of Merkel cell complexes, we measured withdrawal thresholds in mice lacking Merkel cells (*K5Cre*; *Atoh1*$^{flox/flox}$ or *Atoh1*^{cKO}) (*10*–*16*). In mice lacking Merkel cells, Meissner corpuscles were present in slightly increased numbers (Fig. 1C). Behaviorally, mice lacking Merkel cells performed normally on the von Frey paw withdrawal test (Fig. 2B).

Next, we trained water-deprived mice within an operant conditioning behavioral paradigm to report detection of glabrous skin indentation stimuli by lick retrieval of a water reward (Fig. 2, C to E). Mice learned to lick specifically in response to mechanical stimuli within 2 weeks, at which time we delivered step indentations ranging between 0 and 75 mN to the forepaw and measured perceptual detection thresholds (force at the midpoint of the psychometric function; Fig. 2F). Whereas control mice exhibited perceptual detection thresholds of about 10 mN, mice lacking Meissner corpuscles had increased perceptual thresholds of about 20 mN. By contrast, reactivity to a light air puff applied to back hairy skin was comparable in control and *TrkB*^{cKO} mice (fig. S3), indicating normal hairy skin sensitivity.

Meissner corpuscles are required for normal fine sensorimotor behavior

We next tested the requirement of Meissner corpuscles in an innate, fine sensorimotor

[1]Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, MA 02115, USA. [2]Howard Hughes Medical Institute, Harvard Medical School, 220 Longwood Avenue, Boston, MA 02115, USA. [3]Department of Biochemistry, Stanford University, 279 Campus Drive, Stanford, CA 94305, USA. [4]Howard Hughes Medical Institute, Stanford University, 279 Campus Drive, Stanford, CA 94305, USA.
*These authors contributed equally to this work.
†Present address: Two Six Labs, 901 North Stuart Street, Arlington, VA 22203, USA.
‡Corresponding author. Email: david_ginty@hms.harvard.edu

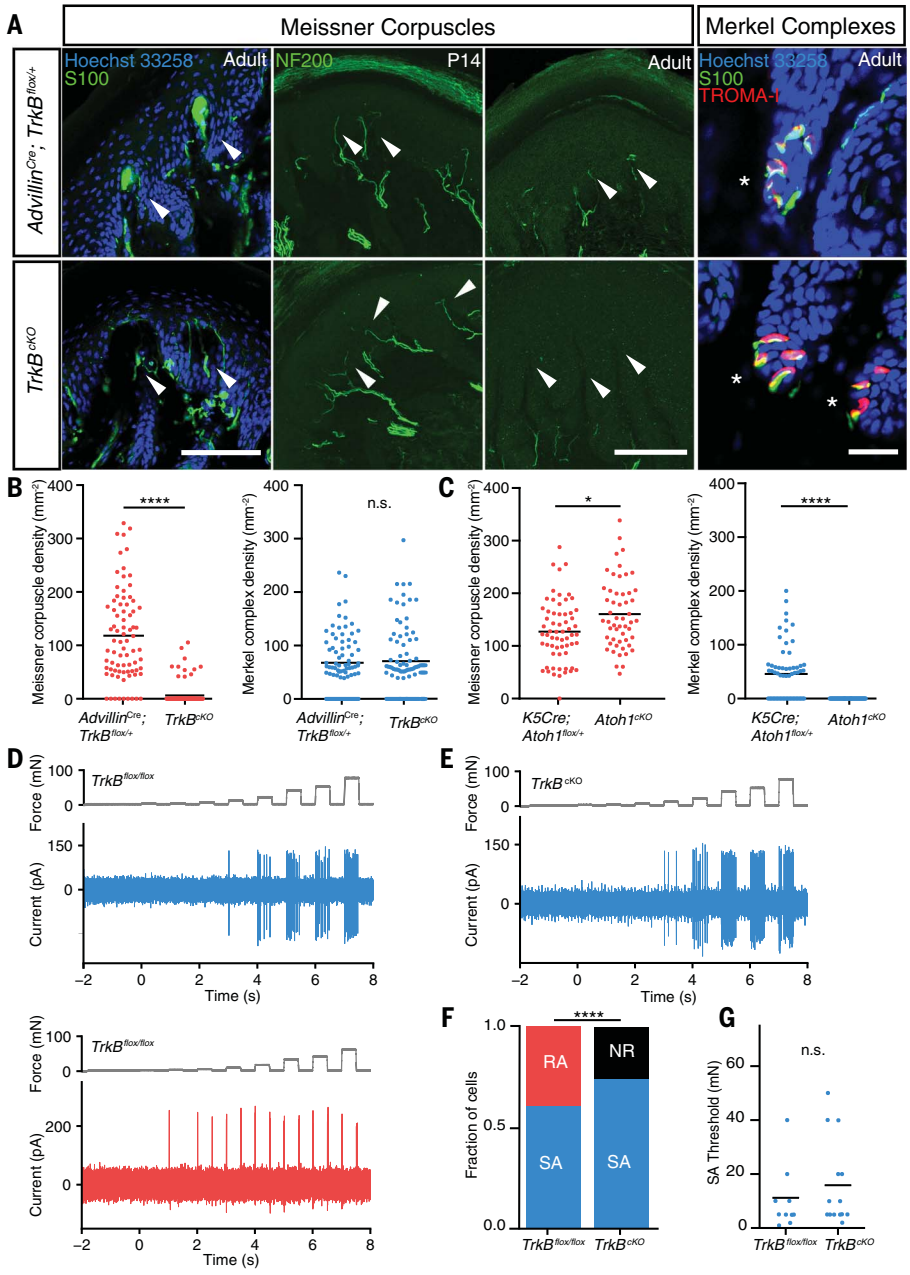


Fig. 1. Meissner corpuscles and their innervating Aβ LTMRs are absent in *TrkB*^{cKO} mice. (**A**) Digital pad sections from control (top) and *TrkB*^{cKO} mice (bottom). Arrows and asterisks indicate typical locations of Meissner corpuscles and Merkel complexes, respectively. Meissner corpuscles and their afferents are labeled with S100 or NFH immunostaining in separate sections, respectively (scale bars, 100 μm). Merkel cells and associated nerve terminals are labeled with TROMA-I and S100, respectively (scale bar, 25 μm). (**B**) Meissner corpuscle and Merkel complex density in glabrous pads of control and *TrkB*^{cKO} mice. Dots represent individual sections, and black bars represent mean [79 sections from three control mice and 81 sections from three *TrkB*^{cKO} mice; two-tailed Mann-Whitney *U* test, ****$p < 0.0001$ ($U = 580$), n.s. indicates not significant ($U = 3163$, $p = 0.9016$)]. (**C**) Meissner corpuscle and Merkel complex density in pads of *K5Cre; Atoh1*$^{flox/+}$ and *Atoh1*^{cKO} (*K5Cre; Atoh1*$^{flox/flox}$) mice. Dots represent densities from individual sections, and black bars represent means [62 sections from two *K5Cre; Atoh1*$^{flox/+}$ mice and 53 sections from two *Atoh1*^{cKO} mice; two-tailed Mann-Whitney *U* test, *$p = 0.0113$ ($U = 1193$), ****$p < 0.0001$ ($U = 689$)]. (**D**) In vivo recordings of a SA (top, blue) and a RA (bottom, red) Aβ LTMR from control mice in response to step indentations applied to glabrous hindpaw pedal pads (force records shown in gray). (**E**) Example recording of an Aβ SA-LTMR from a *TrkB*^{cKO} mouse in response to the same stimulus as in (D). (**F**) Proportion of neurons that are SA, RA, and not responsive (NR) to indentation in control (left, $n = 18$ cells) and *TrkB*^{cKO} mice (right, $n = 20$ cells). χ^2 test: $\chi^2 = 8.5$, $p = 0.004$. (**G**) Mean (black bars) and individual (circles) thresholds of Aβ SA-LTMRs in control (left, $n = 11$ cells) and *TrkB*^{cKO} mice (right, $n = 15$ cells). Unpaired Student's *t* test: $t = -0.81$, $p = 0.42$.

behavior: forepaw manipulation of sunflower seeds while eating. The amount of time required to deshell and eat the sunflower seed kernel was comparable for control and *TrkB*^{cKO} mice. However, whereas control mice used their forepaws to hold, elevate, and rotate the sunflower seed, *TrkB*^{cKO} mice typically trapped or braced the seed on the floor while biting and removing the shell to expose the kernel (Fig. 2G and movies S1 and S2). *TrkB*^{cKO} mice exhibited fewer forepaw dips to the floor and seed rotations and instead more frequently touched the floor to brace the seed (Fig. 2G). By contrast, gait was largely unaltered in *TrkB*^{cKO} mice, although subtle differences in gait were apparent at high velocities (fig. S4). We observed no other obvious sensorimotor deficits in *TrkB*^{cKO} mice (figs. S3 and S4).

Meissner corpuscles are dually innervated

Large-diameter, myelinated sensory neurons innervating Meissner corpuscles express the tyrosine kinase Ret beginning on embryonic day 10.5 (E10.5) (*9*). Consistent with this, *Ret*^{CreER}; *Rosa26*$^{LSL\text{-}tdTomato}$ mice (*9*, *17*) treated with tamoxifen from E10.5 to E13.5 had tdTomato$^+$ sensory fibers that innervate Meissner corpuscles (Fig. 3A). Because TrkB$^+$ sensory neurons are required for Meissner corpuscle formation, we asked whether these Ret$^+$ Meissner afferent fibers also express TrkB. TrkB$^+$ axon terminals were present in the dermal papillae of glabrous skin during embryonic and early postnatal ages, but TrkB expression was undetectable in Meissner afferent terminals in adult mice (fig. S5). Therefore, to permanently label TrkB$^+$ neurons and visualize their cutaneous endings in glabrous skin of adults, we treated *TrkB*^{CreER}; *Rosa26*$^{LSL\text{-}tdTomato}$ mice (*18*) with tamoxifen between E15.5 and P6. We observed that adult *TrkB*^{CreER}; *Rosa26*$^{LSL\text{-}tdTomato}$ mice possess tdTomato$^+$ sensory fibers within Meissner corpuscles (Fig. 3B). These TrkB$^+$ Meissner corpuscle afferents are distinct from Ret$^+$ Meissner corpuscle afferents. Indeed, using *TrkB*^{CreER}; *Rosa26*$^{LSL\text{-}tdTomato}$; *Ret*^{CFP} compound mice (*19*), the TrkB$^+$ and Ret$^+$ sensory neurons were found to be separate populations, with endings of both of these neuronal populations terminating within the majority of corpuscles in both forepaw and hindpaw glabrous skin (Fig. 3, C and E). Dual innervation of corpuscles was also seen using the *Npy2r-GFP* BAC transgenic mouse line, which labels Aβ RA-LTMRs in both hairy and glabrous skin (*20*) (GFP, green fluorescent protein). In *TrkB*^{CreER}; *Rosa26*$^{LSL\text{-}tdTomato}$; *Npy2r-GFP* compound mice, all GFP$^+$ fibers in the dermal papillae were tdTomato$^+$ (fig. S5). On the other hand, in *Ret*^{CreER}; *Rosa26*$^{LSL\text{-}tdTomato}$; *Npy2r-GFP* compound mice, tdTomato$^+$ and GFP$^+$ fibers in dermal papillae were distinct and intertwined (Fig. 3D). Both the Ret$^+$ and TrkB$^+$ Meissner afferents expressed neurofilament, and together

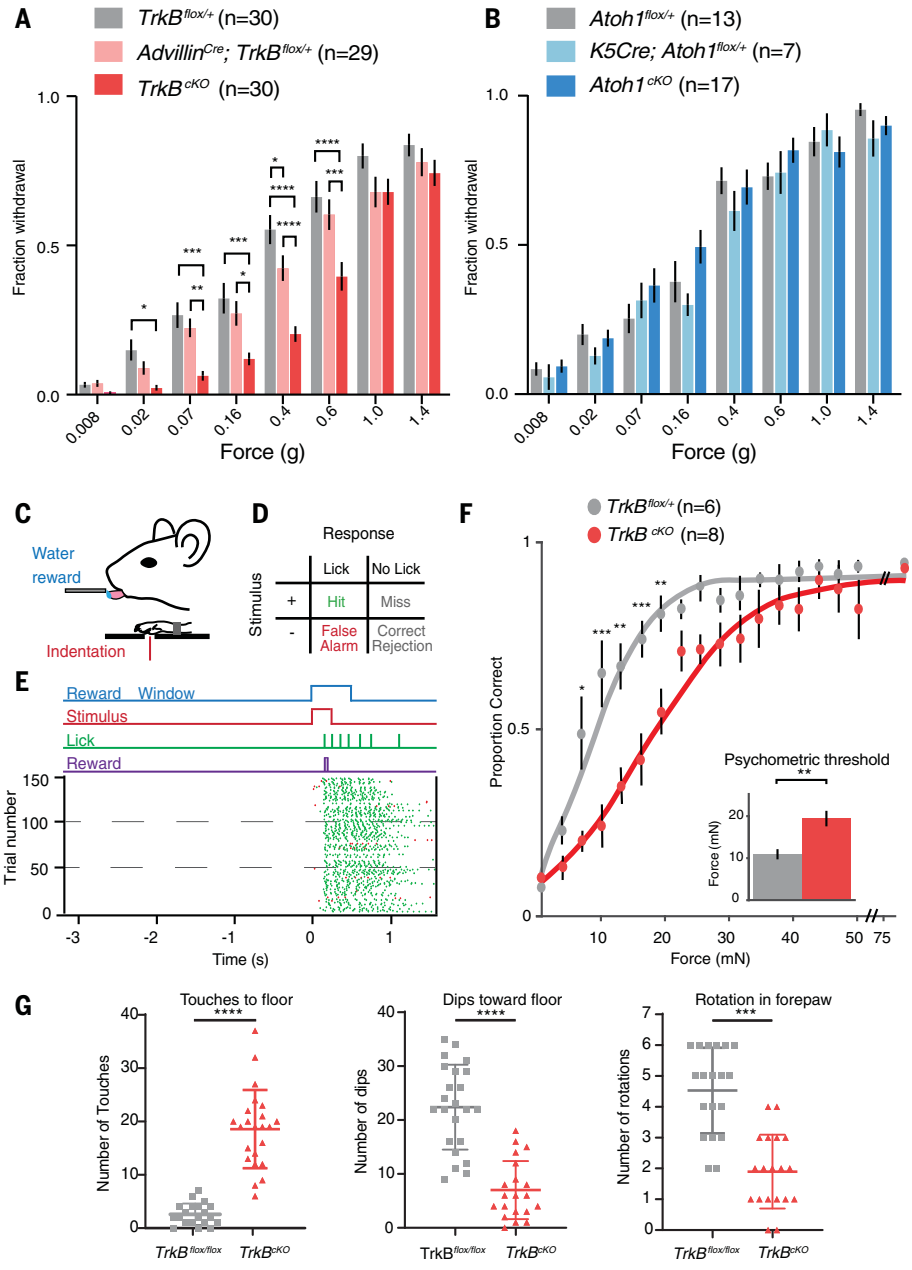


Fig. 2. Meissner corpuscles are necessary for light touch perception and fine sensorimotor control. (**A**) Fraction (mean ± SEM) of paw withdrawals to von Frey filament applications for *TrkB*^{cKO} and control mice. Repeated measures (RM) two-way analysis of variance (ANOVA), effect of genotype ($F_{2,86}$ = 9.823, p = 0.0001) with post hoc Tukey's test; p values represent comparisons between genotypes for each filament: *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. (**B**) Same as (A) for *Atoh1*^{cKO} and control mice. RM two-way ANOVA, no effect of genotype: $F_{2,34}$ = 0.8258, p = 0.4465. (**C**) Operant conditioning task design. (**D**) Matrix of possible behavioral outcomes from one trial. (**E**) Operant conditioning task. The stimulus paradigm, in which animals must withhold for at least 3 s before initiation of a new trial, is depicted (top). In trials with a stimulus (shown), animals receive rewards only on hit trials. Raster plot of licks is shown (bottom). Hit trials are shown with green ticks; false alarm trials are shown with red ticks. (**F**) Psychometric functions for the operant conditioning task for *TrkB*^{cKO} and control mice. Error bars represent SEM (the inset shows thresholds; unpaired Student's t test: t = 2.69, p = 0.002). Two-way ANOVA, effect of genotype ($F_{1,12}$ = 8.261, p = 0.0140) with post hoc Sidak's multiple comparison text; p values represent comparisons between genotypes for each force range: *p < 0.05, **p < 0.01, and ***p < 0.001. (**G**) The number of sunflower seed touch-taps when the seed is braced against the floor between forepaws during seed peeling is shown on the left. The number of times the mouse held the seed in an elevated position and used its incisors to bite into the seed while applying downward force (dips) to expose seed kernel is shown in the middle. The number of seed rotations as mice adjusted their grasp is shown on the right. Unpaired Student's t test, ***p < 0.001 and ****p < 0.0001.

these two neuronal populations accounted for most, if not all, NFH$^+$ Meissner corpuscle endings (Fig. 3C and fig. S5). Meissner corpuscle lamellar cells, which are specialized terminal Schwann cells, migrated into dermal papillae along Meissner afferent axons during development (fig. S6), and TrkB$^+$ axons and corpuscle lamellar cells were both absent from dermal papillae of *TrkB*^{cKO} mice. Although axons of Ret$^+$ neurons still projected into dermal papillae in the absence of TrkB$^+$ axons and corpuscle lamellar cells in neonatal *TrkB*^{cKO} mutants, these Ret$^+$ fibers either retracted or degenerated because neither Ret$^+$ nor TrkB$^+$ afferents were present in dermal papillae of adult *TrkB*^{cKO} mice (fig. S2). Thus, consistent with prior work suggesting multiple, immunohistochemically distinct Meissner corpuscle afferent endings (*21*), our findings reveal that Meissner corpuscles are co-innervated by a pair of intertwined, molecularly and developmentally distinct populations of sensory neurons, one expressing TrkB and another expressing Ret.

Meissner afferents exhibit distinct response properties

Meissner corpuscles are believed to be innervated by one physiological type of sensory neuron, the Aβ RAI-LTMR (*21*–*24*). Our electrophysiological recordings showed that Aβ RA-LTMR responses to glabrous skin stimulation are lost in *TrkB*^{cKO} mice (Fig. 1F). Therefore, we hypothesized that the Ret$^+$ and TrkB$^+$ Meissner afferents are previously unrecognized subsets of Aβ RAI-LTMRs. Aβ RA-LTMRs form central collaterals along the rostrocaudal axis of the spinal cord and also extend an axonal branch via the dorsal column to the dorsal column nuclei (DCN) of the brainstem (*25*–*27*). This projection via the direct dorsal column pathway is likely important for tactile perception and discriminative touch (*21*). We visualized the central projections of both the Ret$^+$ and TrkB$^+$ Meissner afferents. These analyses revealed that both Meissner afferent subtypes display archetypal Aβ LTMR central collateral morphologies and axonal projections to the DCN (fig. S7).

We next asked whether either or both Ret$^+$ and TrkB$^+$ Meissner afferents have physiological response properties consistent with those of Aβ RAI-LTMRs. We used the in vivo DRG recording preparation for targeted electrophysiological recordings of genetically labeled Ret$^+$ and TrkB$^+$ Meissner afferent populations. All recorded Meissner afferents of both subtypes had small, well-defined receptive fields confined to a single glabrous pad. Both subtypes also had conduction velocities within the Aβ range, although the mean conduction velocity of TrkB$^+$ Meissner afferents was modestly but significantly slower than that of Ret$^+$ Meissner afferents (Fig. 4A). TrkB$^+$ Meissner

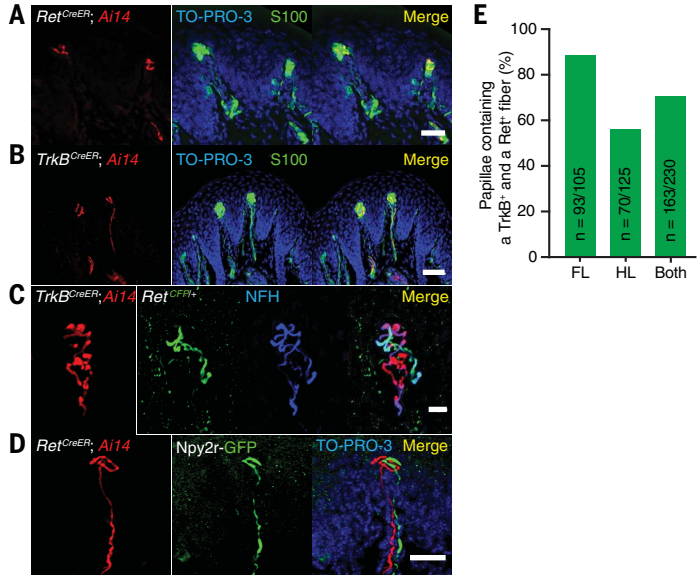


Fig. 3. Meissner corpuscles are innervated by two molecularly distinct mechanosensory neuron types. (**A**) Forelimb pedal pad section of a P20 *Ret*^{CreER}; *Rosa26*$^{LSL\text{-}tdTomato}$ (*Ai14*) mouse treated with tamoxifen at E10.5 to E11.5. The section is stained for S100, DsRed, and TO-PRO-3. This experiment was repeated in three mice. Scale bar, 25 μm. (**B**) Hindlimb digital pad section of a P50 *TrkB*^{CreER}; *Rosa26*$^{LSL\text{-}tdTomato}$ (*Ai14*) mouse treated with tamoxifen at E16.5. The section is stained for S100, DsRed, and TO-PRO-3. This experiment was performed in eight mice with varying dates of tamoxifen administration (E13.5, E16.5, P2, and P6). Scale bar, 50 μm. (**C**) Hindlimb digital pad section of a P50 *TrkB*^{CreER}; *Ai14*; *Ret*^{CFP} mouse treated with tamoxifen at P5. The section is stained for DsRed, GFP, and NFH. This experiment was repeated in four mice with tamoxifen administration at E13.5 or P5. Scale bar, 10 μm. (**D**) Forelimb pedal pad section of a P20 *Ret*^{CreER}; *Ai14*; *Npy2r-GFP* mouse treated with tamoxifen at E10.5. The section is stained with anti-DsRed, anti-GFP, and TO-PRO-3. This experiment was repeated in three mice. Scale bar, 25 μm. (**E**) Percentage of dermal papillae containing both TrkB$^+$ and Ret$^+$ Meissner afferents in forelimb pads, hindlimb pads, and all pads, as measured in *TrkB*^{CreER}; *Ai14*; *Ret*^{CFP} mice (two animals). Only papillae containing a tdTomato$^+$ fiber were included in the quantification to eliminate an effect of Cre-lox efficiency on the interpretation of these measurements.

afferents responded at both the onset and offset of force-controlled step indentations with high sensitivity, whereas Ret$^+$ Meissner afferents responded at the onset of step indentations with lower and more varied sensitivity and rarely responded to the offset of step indentations (Fig. 4, B to D). Some (4 of 10) of the Ret$^+$ Meissner afferents responded during the sustained phase of the step indentation (example in Fig. 4E), indicating that Meissner-innervating afferents are not always rapidly adapting. Both TrkB$^+$ and Ret$^+$ Meissner afferents exhibited a wide range of frequency tuning to 2- to 120-Hz sinusoidal vibrations: Some neurons of both types responded best to high frequencies, whereas others had nearly uniform force thresholds across the tested frequencies (fig. S8). Although frequency tuning did not differ between TrkB$^+$ and Ret$^+$ Meissner afferents, TrkB$^+$ Meissner afferents were more sensitive to vibratory stimuli than Ret$^+$ Meissner afferents (fig. S8). Altogether, although the response properties of TrkB$^+$ Meissner afferents fit the classic Aβ RAI-LTMR definition, those of most Ret$^+$ Meissner afferents do not.

Ret$^+$ and TrkB$^+$ Meissner afferents are homotypically tiled and heterotypically offset

We next investigated the spatial relationships between the cutaneous termination fields of Ret$^+$ and TrkB$^+$ Meissner afferents. Homotypic tiling is a phenomenon in which the anatomical receptive fields of neurons do not spatially overlap with other neurons of their same type. Using whole-mount alkaline phosphatase (AP) staining of sparsely labeled Meissner afferents, we found that the skin innervation areas of individual Ret$^+$ and TrkB$^+$ Meissner afferents are well demarcated, spatially contiguous, and confined to single glabrous pads (fig. S9), consistent with the physiological receptive fields being confined to single glabrous pads. Occasionally, two neurons innervating the same toe or pedal pad were labeled (Fig. 5A). In these cases, TrkB$^+$ afferents were never observed to spatially overlap, and only 1 out of 10 pairs of Ret$^+$ afferents occupying the same pad displayed spatial overlap (Fig. 5C). For the single overlapped Ret$^+$ afferent pair, the two arborizations shared only 1 out of 12 total corpuscles, and this corpuscle was on the border of their termination fields. We also used an intersectional genetic labeling strategy to address homotypic tiling. We generated mice harboring either the *TrkB*^{CreER} or *Ret*^{CreER} allele in conjunction with two different Cre-dependent fluorescent reporters, *Rosa26*$^{LSL\text{-}YFP}$ and *Rosa26*$^{LSL\text{-}tdTomato}$. We treated these mice with high doses of tamoxifen to randomly and efficiently label individual afferents with one of three possibilities: tdTomato, yellow fluorescent protein (YFP), or both. Virtually all Meissner corpuscles examined in these mice contained a single-color fiber of a given afferent subtype (Fig. 5, B and D), thus confirming homotypic tiling of the two Meissner afferent types. We also compared the peripheral arborizations of sparsely labeled Ret$^+$ and TrkB$^+$ Meissner corpuscle afferents. These two populations exhibited differences in both their surface-area size distributions and the number of Meissner corpuscles innervated by individual neurons (Fig. 5, E and F, and fig. S9), indicating that the two homotypically tiled Meissner afferent types are heterotypically offset.

One potential advantage of heterotypic overlap of Meissner mechanoreceptors with different force sensitivities is to enable discrimination of a greater range of forces. In this model, spikes emanating from the TrkB$^+$ Meissner Aβ LTMR alone encode the lightest indentation forces, whereas spikes from both TrkB$^+$ and Ret$^+$ Meissner mechanoreceptors encode higher forces. Another potential advantage of heterotypic overlap of homotypically tiled Meissner mechanoreceptors is that population responses to forces that coactivate both Meissner afferent types could enable a heightened ability to detect and localize a point stimulus. We modeled the receptive fields of each of the two Meissner afferent subtypes as homotypically tiled grids that are heterotypically offset and simulated measurements of spatial acuity (fig. S10). Our simulations indicated that for a single nonoverlapping grid, the number of discriminable spatial locations is equal to the number of neurons (fig. S10). Simulations using two grids that are homotypically tiled but overlapping and maximally offset with respect to each other revealed that the number of distinct spatial locations doubles for the same number of neurons (fig. S10). This gain arose from a distributed neural representation that an overlapped arrangement facilitates and may enable an enhanced ability to localize a point stimulus without increasing the number of neurons. Therefore, although functional analyses of mice selectively lacking one or the other Meissner afferent subtype will be needed to test our mathematical modeling findings, we speculate that the heterotypic overlap of homotypically tiled Meissner mechanoreceptor subtypes may enable heightened discriminatory capabilities, at a population level, compared with a simpler arrangement of a single homotypically tiled Meissner afferent type.

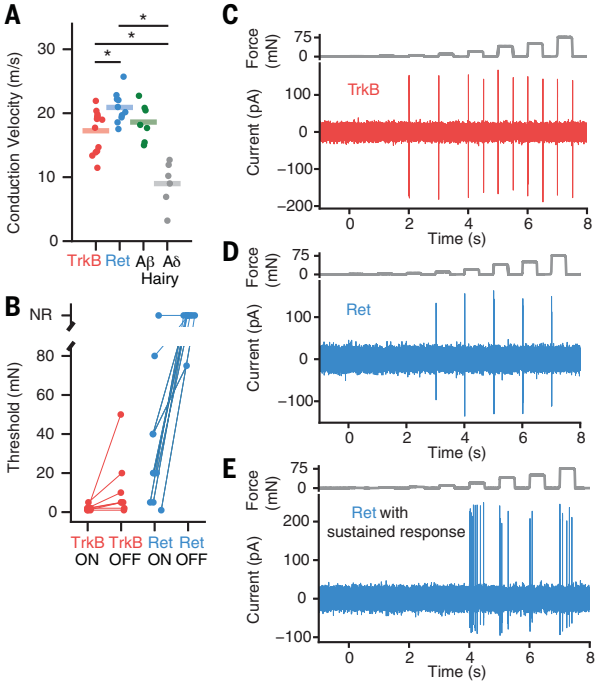


Fig. 4. Ret$^+$ and TrkB$^+$ Meissner afferents exhibit distinct physiological response properties. (**A**) Both Meissner afferent subtypes have conduction velocities in the Aβ range as defined by the range of conduction velocities measured from hairy skin Aδ and Aβ LTMRs in the same preparation. The average conduction velocities for TrkB$^+$ and Ret$^+$ Meissner afferents are significantly different [10 afferents per type; mean ± SEM: TrkB = 17.3 ± 0.9 m/s and Ret = 20.9 ± 0.8 m/s; Mann-Whitney *U* test (U = 26.0, p = 0.008)]. *p < 0.05. (**B**) The minimal force required to produce an action potential at the onset or offset of a step indentation for TrkB$^+$ and Ret$^+$ Meissner afferents. NR, not responsive. (**C**) Response of a TrkB$^+$ Meissner Aβ LTMR to a series of step indentations increasing in intensity from 1 to 75 mN. (**D**) Response of a RA Ret$^+$ Meissner afferent to the same stimuli as in (C). (**E**) Response of a SA Ret$^+$ Meissner afferent to the same stimuli as in (C).

Ret$^+$ and TrkB$^+$ Meissner afferents have distinct lamellar cell associations

The different sensitivities and response kinetics of Ret$^+$ and TrkB$^+$ Meissner Aβ mechanoreceptors raised the question of whether these mechanoreceptor types form distinct endings within the corpuscle where mechanotransduction occurs. We evaluated the ultrastructural properties of the axonal endings of Ret$^+$ and TrkB$^+$ Aβ mechanoreceptors within the Meissner corpuscle using electron microscopy (EM). We crossed *Ret*^{CreER} or *TrkB*^{CreER} mice to reporter mice that express the peroxidase dAPEX2 targeted to the mitochondrial matrix in a recombinase-dependent manner (*28*), thereby generating an electron-dense label associated with mitochondria and distributed throughout the labeled cells. We could distinguish and characterize the ultrastructural properties of labeled Aβ mechanoreceptors within the Meissner corpuscle. Axons of both Ret$^+$ and TrkB$^+$ mechanoreceptor subtypes terminated within the Meissner corpuscle, and both ending types were intimately associated with resident, morphologically complex corpuscle lamellar cells (Fig. 6, A and B). The lamellar cell processes that surround both Meissner mechanoreceptor ending types were periodically discontinuous, thus exposing the naked mechanoreceptor axonal membrane to the extracellular matrix (Fig. 6C). Whereas Ret$^+$ Meissner axon profiles were usually surrounded by only a few concentrically arranged lamellar cell processes, the lamellar cell processes associated with TrkB$^+$ Meissner profiles were typically much more elaborate. Indeed, TrkB$^+$ Meissner afferents have a minimum of five concentric lamellar processes surrounding the axon terminal (range 5 to 28) and, on average, twice as many processes compared with Ret$^+$ Meissner afferent terminals (range 3 to 18; Fig. 6D). This ultrastructural difference is particularly noteworthy when considering the distinct response kinetics of Ret$^+$ and TrkB$^+$ Meissner afferents together with classical studies of the Pacinian corpuscle suggesting that lamellar wrappings are critical for the Pacinian LTMR's response kinetics (*29*). For the Pacinian afferent, lamellar cell wrappings are believed to be essential for its high selectivity to vibratory stimuli and for producing a generator potential as force is withdrawn during a step offset. Moreover, "unwrapping" or removing Pacinian lamellar cells converted its generator potential from being transient to sustained (*29*). Our findings that Ret$^+$ Meissner mechanoreceptors are associated with fewer lamellar cell wrappings compared with the TrkB$^+$ Meissner LTMR and that they lack responses at indentation step offsets and occasionally display slowly adapting responses are consistent with these classical theories.

Conclusions

Our findings show the Meissner corpuscle to be crucial for perception of the gentlest detectable forces acting on the glabrous skin and for fine sensorimotor control during object manipulation. The corpuscle is endowed with intertwined endings of two molecularly, ultrastructurally, and physiologically distinct mechanoreceptor subtypes whose cutaneous innervation areas are homotypically tiled and heterotypically offset. This arrangement of differentially sensitive Meissner Aβ mechanoreceptors supports a "population coding" model for force intensity. Computational simulations suggest that it may enable higher acuity, at a population level, than would be expected from a simpler arrangement with a single Meissner mechanoreceptor subtype. The organization of the Meissner corpuscle afferents' termination fields thus illustrates a solution to the nervous system's general challenge of localizing stimuli with a limited number of neurons. Our EM findings reveal that the axonal endings of the Ret$^+$ and TrkB$^+$ Meissner Aβ mechanoreceptor subtypes, and in particular their association with concentrically organized processes of corpuscle lamellar cells, are distinct, supporting a model in which ultrastructural properties of mechanoreceptors within a corpuscle dictate their responses.

Materials and methods

Animals

Animals were handled according to protocols approved by the Harvard Standing Committee on Animal Care and are in accordance with federal guidelines.

Mouse lines

The *TrkB*^{flox}, *Advillin*^{Cre} (JAX 032536), *K5Cre* (MGI 3050065), *Atoh1*^{flox} (JAX 008681), *Ret*^{CreER} (MGI #4437245), *Ai14* (*Rosa26*$^{LSL\text{-}tdTomato}$; JAX 007914), *Ret*^{CFP} (MGI 3777555), *Npy2r-GFP* (MGI 3844094), *Rosa26*^{iAP} (JAX 009253), *Ai3* (*Rosa26*$^{LSL\text{-}YFP}$; JAX 007903), *Ai32* (*Rosa26*$^{LSL\text{-}ChR2(H134R)\text{-}EYFP}$; JAX 024109), *TrkB*^{CreER} (JAX 027214), *Pvalb*$^{2A\text{-}FlpO\text{-}D}$ (JAX 022730), *Ai65* (*Rosa26*$^{FSF\text{-}LSL\text{-}tdTomato}$; JAX 021875), *Rosa26*$^{LSL\text{-}FSF\text{-}ReaChR\text{-}mCitrine}$ (JAX 024846), *Rosa26*$^{DR\text{-}Matrix\text{-}dAPEX2}$ (JAX 032764), *Rosa26*$^{LSL\text{-}Matrix\text{-}dAPEX2}$ (JAX 032765), *Brn3a*$^{f(AP)}$ (JAX 010558), *BDNF*^{lacZ} (*BDNF*^{flox}; JAX 021055), *Wnt1Cre* (JAX 022137), and *DhhCre* (JAX 012929) mouse lines have been described previously (*5*, *6*, *9*, *14*, *15*, *17*–*20*, *28*, *30*–*38*). The *Advillin*^{FlpO} mouse line, which was generated by standard gene targeting procedures, enables FlpO-dependent recombination in somatosensory and other peripheral nervous system neurons and will be described in detail elsewhere. All lines were kept on a mixed CD1-C57Bl/6 background.

Tamoxifen treatments

Tamoxifen (Toronto Research Chemicals) was dissolved in 100% ethanol (20 mg/ml). The tamoxifen-ethanol mixture was diluted with sunflower seed oil (Sigma) at a 1:2 to 1:3 dilution and vacuum centrifuged for 30 to 45 min until the ethanol evaporated. The resulting solution was stored at −20°C and thawed immediately before use. For embryonic tamoxifen treatment, the oil-tamoxifen solution was delivered to pregnant mothers by oral gavage. The

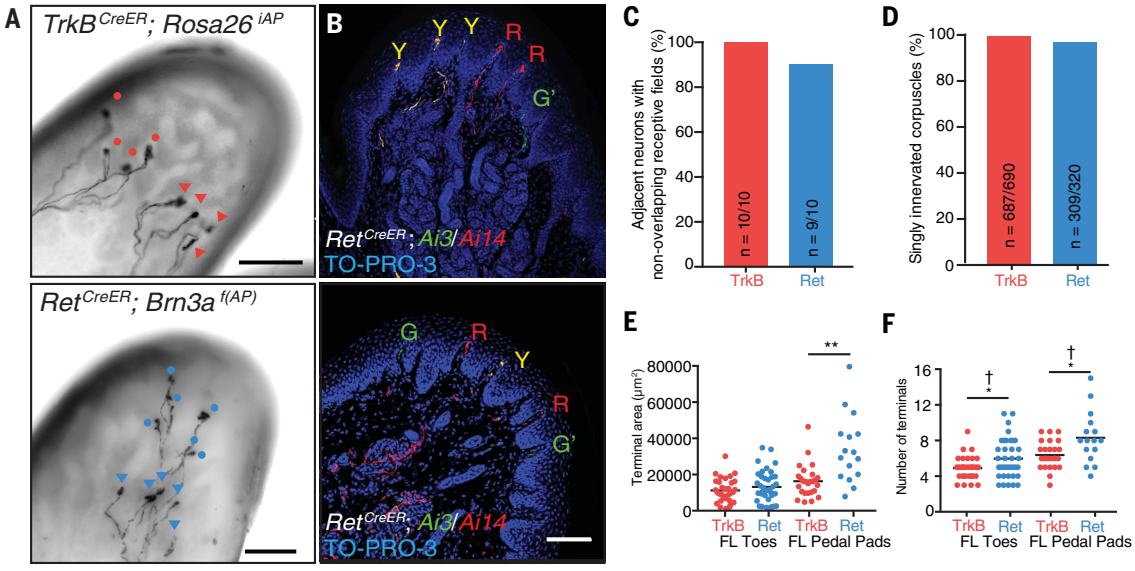


Fig. 5. Spatial arrangement of Ret$^+$ and TrkB$^+$ Meissner afferent cutaneous endings. (**A**) Images of whole-mount, AP-stained digital pads of a *TrkB*^{CreER}; *Rosa26*^{iAP} mouse (top) and a *Ret*^{CreER}; *Brn3a*$^{f(AP)}$ mouse (bottom, Z-projection performed in Fiji), depicting pairs of afferents of the same subtype innervating the same glabrous pad. Tamoxifen doses were titrated to achieve sparse labeling of Meissner afferents. Terminals of one neuron are annotated with circles and terminals of the other with triangles. Scale bars, 100 μm. (**B**) Digital pad sections of *TrkB*^{CreER}; *Rosa26*$^{LSL\text{-}YFP/LSL\text{-}tdTomato}$ (*Ai3/Ai14*) mice treated with tamoxifen at E12.5 and E13.5 (top) and *Ret*^{CreER}; *Ai3/Ai14* mice administered tamoxifen at E11.5 and E12.5 (bottom). Y, fibers express both tdTomato and YFP; R, fibers express tdTomato only; G, fibers express YFP only; G′, YFP$^+$ fibers without terminals in the sections. Sections were stained with anti-DsRed, anti-GFP, and TO-PRO-3. Scale bar, 100 μm. (**C**) Percentage of TrkB$^+$ and Ret$^+$ Meissner afferent pairs innervating the same pad and occupying different spatial territories. (**D**) Quantification of the percentage of dermal papillae and Meissner corpuscles receiving a single TrkB$^+$ fiber (three animals) and a single Ret$^+$ fiber (six animals). (**E** and **F**) Receptive field surface area [terminal area, (E)] and number of terminal endings (F) of TrkB$^+$ and Ret$^+$ Meissner afferents measured in *TrkB*^{CreER}; *Rosa26*^{iAP} or *TrkB*^{CreER}; *Brn3a*$^{f(AP)}$ mice [58 forelimb (FL) receptive fields from 21 mice] and *Ret*^{CreER}; *Rosa26*^{iAP} or *Ret*^{CreER}; *Brn3a*$^{f(AP)}$ mice (53 FL receptive fields from 21 mice). Individual measurements and mean values (black bar) are plotted for forelimb digital pads and pedal pads (two-tailed Welch's *t* test, mean significantly different: $*p < 0.05$ and $**p < 0.01$; *F* test of the equality of variances, variance significantly different: $\dagger p < 0.01$).

Ret$^+$ Meissner afferent population was labeled with high efficiency by giving pregnant mothers a 2-mg dose of tamoxifen at E11.5 and E12.5. The forelimb was labeled more effectively at E11.5, whereas the hindlimb was labeled more effectively at E12.5. The TrkB$^+$ Meissner afferent population was labeled with high efficiency by intraperitoneal injection at P4, but this population was also labeled when pregnant mothers were given tamoxifen any time between E12.5 and E16.5 (2 mg) and when P2 to P6 pups were given tamoxifen (0.04 to 0.1 mg/g body weight).

Immunohistochemistry

P20 or older mice were euthanized with CO_2 and perfused with Ames Media (Sigma) containing 10 U/ml heparin (Sigma) in phosphate-buffered saline (PBS), followed by 4% paraformaldehyde (PFA) in PBS. Spinal cords and brainstems were dissected and fixed in 4% PFA at 4°C overnight. For skin sectioning experiments, glabrous toe and pedal pads were removed from the paw and placed in 4% PFA for 4 to 6 hours or Zamboni's fixation buffer for 24 to 48 hours at 4°C. For whole-mount immunohistochemistry, whole paws were placed in Zamboni's fixation buffer at 4°C for 24 to 48 hours. The glabrous skin was then separated from the underlying tissue by fine dissection in PBS. Isolated glabrous paws were fixed in Zamboni's fixation buffer for an additional 1 hour at room temperature. Following fixation, all tissues were rinsed three times for 10 min in PBS at room temperature. Brainstems and spinal cords with dorsal root ganglia attached were isolated from the vertebral column.

Glabrous pads, spinal cords, and brainstems destined for sectioning were cryoprotected in 30% sucrose in PBS at 4°C for 24 to 48 hours. Tissues were embedded in optimal cutting temperature (OCT) compound (Tissue Tek) and frozen at −20°C. Glabrous pads were sectioned at 25 μm, and spinal cords and brainstems were sectioned transversely at 40 to 50 μm. Sections were dried overnight at room temperature. Glabrous pads were given one to two additional days to dry. All sections were rinsed three times for 10 min with PBS, blocked with 5% normal serum (donkey or goat) in 0.1% PBST (0.1% Triton X-100 in PBS) for 2 hours, and incubated with primary antibody in blocking solution overnight at 4°C. The following day, sections were rinsed three times for 10 min with PBST and incubated with secondary antibodies in blocking solution for 2 hours at room temperature or overnight at 4°C. Sections were then rinsed once for 10 min in PBST and twice for 10 min in PBS and mounted with Fluoromount-G (Southern Biotech). Images were acquired on a Carl Zeiss LSM 700 confocal microscope.

Whole-mount immunohistochemistry of glabrous skin was performed as described previously (7). The skin was given several 1-hour washes in 0.3 to 0.5% TritonX-100 in PBS. The skin was then incubated in primary antibodies in blocking solution [5% normal serum (goat or donkey), 75% PBST, 20% DMSO] at room temperature for 72 hours. The skin was washed in PBST for four to five times for 1 hour and incubated in secondary antibodies in blocking solution at room temperature for 48 hours. Tissue was washed several times in PBST then dehydrated in serial methanol dilutions (1 hour each of 25%, 50%, 75%, and 100%), followed by overnight dehydration in 100% methanol at room temperature. The tissue was then cleared in BABB (1 part benzyl alcohol: 2 parts benzyl benzoate) and imaged.

Primary antibodies were rabbit anti-DsRed (Clontech, cat. no. 632496, 1:500), goat anti-mCherry (Sicgen, cat. no. AB0040-200, 1:500), chicken anti-GFP (Aves Lab, cat. no. GFP-1020, 1:500), rabbit anti-GFP (Invitrogen, cat. no. A-11122, 1:500), goat anti-GFP (US Biological, cat. no. G8965-01E, 1:500), rat anti-MBP (Millipore, cat. no. MAB386, 1:500), rabbit anti-NFH (Sigma, cat. no. N4142, 1:1000), chicken anti-NFH (Aves Lab, cat. no. NFH, 1:500), rabbit anti-S100 (Dako, cat. no. Z031129-2, 1:300), rabbit anti-PKC γ (Santa Cruz, cat. no. SC-211, 1:500), and TROMA-I (rat anti-keratin 8/cytokeratin 18, DSHB, University of Iowa, TROMA-I supernatant, 1:100). The nucleic acid stain TO-PRO-3 (Invitrogen, cat. no. T3605, 1:500), the nonpeptidgeric nociceptor

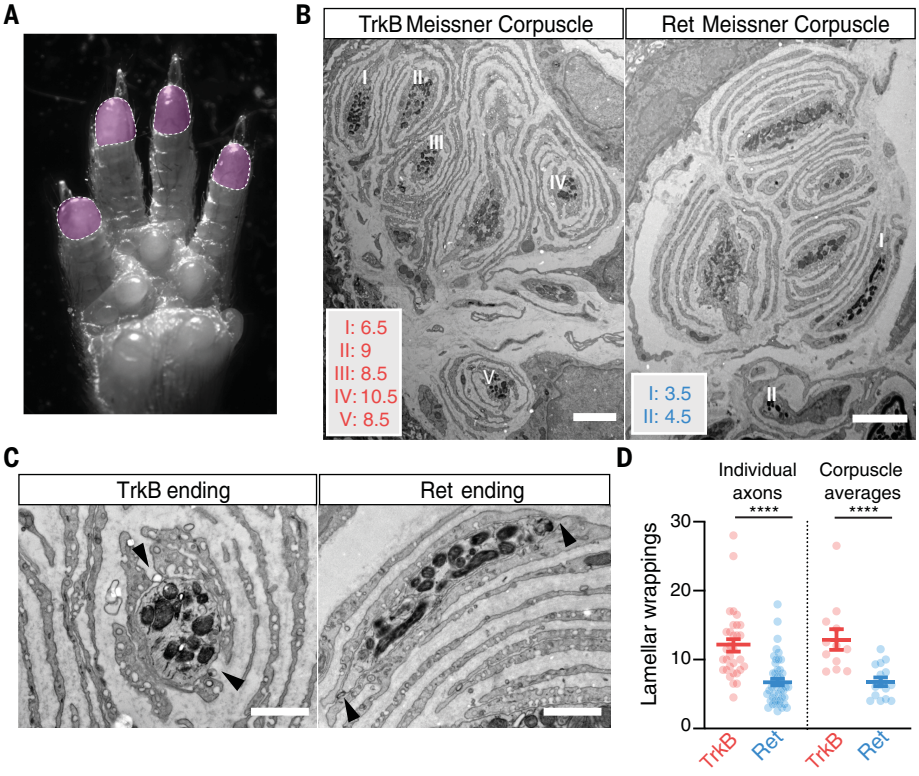


Fig. 6. TrkB$^+$ Meissner Aβ LTMR endings have more lamellar cell wrappings than Ret$^+$ Meissner mechanoreceptor endings. (**A**) Image of a paw, with the shaded areas (digit tips) representing regions used for EM analysis that contain a high density of Meissner corpuscles. (**B**) EM images of Meissner corpuscles from a *TrkB*^{CreER}; *Advillin*^{FlpO}; *Rosa26*$^{DR\text{-}Matrix\text{-}dAPEX2}$ mouse treated with tamoxifen at P3 (left) and a *Ret*^{CreER}; *Advillin*^{FlpO}; *Rosa26*$^{DR\text{-}Matrix\text{-}dAPEX2}$ mouse treated with tamoxifen at E11.5 and P10 (right). Roman numerals mark labeled TrkB (left) and Ret (right) axons. The number of lamellar wrappings for individual labeled axonal profiles is shown. Scale bars, 3 μm. (**C**) High-magnification images of labeled endings and associated lamellar cells from (B). Both exhibit openings or areas not directly associated with lamellar cells (black arrowheads). Scale bars, 1 μm. (**D**) Quantification of the number of lamellar cell wrappings around genetically labeled endings of TrkB and Ret endings. Shown are the number of lamellar wrappings for individual axonal profiles (left; n = 32 and 47, respectively) and averages of all axonal profiles within individual corpuscles (right; n = 12 and 15, respectively). Mann-Whitney U test, ****$p < 0.0001$. Corpuscles imaged from two mice for each LTMR subtype.

marker IB4 (Isolectin GS-IB$_4$, conjugated to Alexa Fluor 647, Life Technologies, cat. no. I32450, 1:500), and the nucleic acid stain Hoechst 33258 (Fisher, cat. no. H3569, 1:1000) were applied during secondary antibody incubation.

X-gal staining

Embryos and glabrous skin were fixed with a glutaraldehyde solution (0.2% glutaraldehyde, 2 mM $MgCl_2$ in PBS) overnight at 4°C. For whole-mount staining, fixed embryos or skin were washed with detergent rinse buffer (0.01% sodium deoxycholate, 0.02% NP40, 2 mM $MgCl_2$ in phosphate buffer, pH 7.4) and stained with staining buffer (0.01% sodium deoxycholate, 0.02% NP40, 2 mM $MgCl_2$, 5mM potassium ferricyanide, 5mM potassium ferrocyanide, 1mg/ml 5-bromo-4-chloro-indolyl-β-D-galactopyranoside in phosphate buffer, pH 7.4) at room temperature. After staining, embryos were dehydrated sequentially with 50% methanol, 80% methanol, and 100% methanol. For staining on sections, 12-μm sections were cut and dried for several hours at room temperature. The staining procedure was the same as described for whole-mount staining. After staining, sections were fixed overnight at 4°C in 4% PFA in PBS, and mounted with Fluoromount-G.

Fluorescent sparse labeling

TrkB$^+$ Meissner afferents were sparsely labeled by injecting the pedal pads of *TrkB*^{CreER}; *Ai65* mice with a viral vector containing FlpO recombinase, AAV1-hSyn-FlpO (AAV2/1.humanSynapsin1.FlpO.bGH, UPenn Vector Core, titer = 1.23×10^{13} genome copies/ml). Mice were administered 0.5 mg of tamoxifen at P5 by interperitoneal injection. Mice aged P9 were anesthetized with isoflurane (2 to 3%) delivered from a precision vaporizer. Animals were monitored throughout the 5- to 10-min procedure, and the anesthetic dose was adjusted as necessary. AAV1-hSyn-FlpO combined with a small amount of Fast Green dye (Sigma F7252-5G) in 0.9% saline was injected into one to three hindlimb or forelimb pedal pads per animal with a beveled glass capillary needle (FHC Inc capillary tubing, FHC 27-30-0). The total volume of AAV–Fast Green mixture injected per pad for sparse labeling was about 0.1 to 0.5 μl. Mice were removed from anesthesia, administered analgesic (Carprofen, 4 mg/kg), and placed on a warm pad for recovery. Animals were given an additional dose of Carprofen after 24 hours and observed to ensure adequate recovery. Mice were perfused 2 to 9 weeks after injection, and their vertebral columns and glabrous paws were prepared as described above. The viral dose was titrated to achieve labeling of a single neuron per animal, and only mice in which a single cervical or lumbar neuron was observed in the DRG were included in subsequent analyses.

Alkaline phosphatase histochemistry

Placental alkaline phosphatase (PLAP) staining was performed as described previously (*6*). Whole paws from P21 mice were fixed in Zamboni's fixation buffer at 4°C for 24 to 48 hours. Age was kept consistent to prevent age and size differences from contributing to observed differences between Ret$^+$ Meissner afferent and TrkB$^+$ Meissner afferent receptive fields. The glabrous skin was then separated from the underlying tissue by fine dissection in PBS. Isolated glabrous paws were fixed in Zamboni's fixation buffer for an additional 1 hour at room temperature. After fixation, all tissues were rinsed three times for 10 min in PBS at room temperature and then incubated at 65° to 68°C for 2 hours. To detect PLAP signal, tissues were incubated with BCIP/NBT (Roche) solution (diluted in 0.1 M Tris pH 9.5, 0.1 M NaCl, 50 mM $MgCl_2$, 0.1% Tween-20 solution) overnight at room temperature. Tissues were then fixed in 4% PFA in PBS for 1 hour at room temperature, followed by dehydration in ethanol (1 hour 50%, 1 hour 75%, 1 hour 100%, overnight 100%). Finally, tissues were cleared in BABB and imaged.

To sparsely label Ret$^+$ Meissner afferents, pregnant mice were administered 1 mg tamoxifen at either E11.5 or E12.5. For TrkB$^+$ Meissner afferents, pregnant mice were administered 1 mg tamoxifen at E14.5.

Quantification of receptive field geometries

Images of Meissner afferent terminal arborizations in glabrous pads were analyzed in the Fiji distribution of ImageJ (www.fiji.sc). The perimeter of individual receptive fields (terminal area) was outlined by hand, and the resulting selection was subjected to the Convex Hull function, which replaces a polygon with its convex hull. The area in μm^2 and the aspect ratio (major axis/minor axis) of the convex hull were measured using the measure function. Terminals of afferent arborizations were counted by eye. For some neurons, two fibers from the

same neuron could be seen traveling together and terminating in the same corpuscle. In these cases, the two fibers were counted as a single termination. This phenomenon appeared to occur more often for $TrkB^+$ Meissner afferents than for Ret^+ Meissner afferents, but this was not quantified.

In vivo electrophysiological recordings

In vivo DRG recordings were made as described previously (*7*) using a preparation modified from that published by Ma, Donnelly, and LaMotte (*8*). *Ret*^{CreER} or *TrkB*^{CreER} mice with a Cre-dependent YFP, tdTomato, or mCitrine reporter (*Ai14*, *Ai32*, or *Rosa26*$^{LSL\text{-}ReaChR\text{-}mCitrine}$) were administered tamoxifen at the appropriate ages. Because this strategy also labels hairy-skin-innervating neurons, we retrogradely labeled glabrous neurons by subcutaneous injection of cholera toxin subunit B conjugated to Alexa 555 or Alexa 488 (Thermo Fisher; 2 μg/μl in PBS, about 0.3 μl per pad) into the left hindlimb toe and pedal pads 2 to 3 days before recording. On the day of recording, P20 to P60 mice were anesthetized with urethane (1 g/kg) by intraperitoneal injection. Anesthesia was maintained with isoflurane (1.5 to 2% in 100% O_2) for the duration of the surgery and subsequent recording period (SomnoSuite, Kent Scientific). Mouse internal temperature was monitored with a rectal probe and maintained at 35.5° to 37.5°C with a temperature controller (TC-344B, Warner Instruments) and thermoelectric heater (C3200-6145, Honeywell) embedded in castable cement (Aremco). The lumbar vertebral column was exposed and secured with custom spinal clamps (Mike's Machine, Attleboro, MA). Bone dorsal to the target DRG(s) (L3 to L5) was removed with rongeurs. The surgical site was continuously perfused with and immersed in external solution containing (in mM) 140 NaCl, 3.1 KCl, 0.5 KH_2PO_4, 6 glucose, 1.2 $CaCl_2$, 1.2 $MgSO_4$, and 10 HEPES. pH was adjusted to 7.4 with NaOH. The epineurium surrounding the DRG was removed with fine forceps. The DRG was visualized using custom reflective optics on an upright compound microscope (Zeiss). Cell bodies on the surface of the DRG were accessible for recording. Borosilicate glass pipettes (TW150F-4, WPI) were pulled to achieve a 20- to 30-μm tip diameter. Pipettes were then filled with external solution, and fluorescent cell bodies that were labeled with dye-conjugated CTB were targeted for cell-attached recordings.

Neuron receptive fields (RFs) were localized using a small paintbrush and gentle manual probing. For Meissner afferents, RFs were confined to single toe or pedal pad in the hindlimb. To estimate conduction velocity (CV), an electrical stimulus was delivered to the RF using a bipolar electrode. The delivered current was 2.5 times the threshold at which the neuron fired an action potential. CVs were estimated by dividing the conduction latency (average latency out of 10 trials) by the distance between the DRG and the RF. This distance did not account for the tortuosity of the sensory axon, and thus, the reported CVs are a lower bound on the true value. Force-controlled indentation of skin was achieved using an indenter (Model 300 C-I, Aurora Scientific) mounted on two orthogonal linear motorized stages (MTS25/M-Z8E, Thorlabs). The XY position of the indenter was controlled using the stages. The indenter was positioned on top of the RF such that the tip of the indenter was barely touching the surface of the glabrous pad. Sinusoidal force stimuli and low-pass filtered (15-ms boxcar) force steps were synthesized in MATLAB (Mathworks, Natick, MA) and delivered to the skin.

For random recordings of Aβ LTMRs that innervate glabrous skin (Fig. 1), the hindpaw glabrous skin pedal pads of *TrkB*$^{flox/flox}$ controls and *TrkB*^{cKO} mice were injected with CTB conjugated to Alexa 488 24 to 72 hours before recording. Neurons with large soma diameters (>25 μm) that contained CTB 488 were targeted for recording. The receptive fields of the neurons were localized with manual brushing, and force indentations (0.5-s duration, 1- to 75-mN intensity) were applied to the receptive field. RA neurons were defined as those with spiking responses to indentation at only the onset or offset of the step indentation. SA neurons were defined as those with spikes occurring at the onset and the middle of the step indentation. SA neurons never increased their firing rate in response to the termination of the indentation. Nonresponsive (NR) neurons were defined as those that produced action potentials in response to stretching of the foot, but not to the indentation. Extracellular action potentials were recorded using a Multiclamp 700A amplifier (Axon Instruments) operating in the voltage-clamp configuration, which was constantly adjusted so that no current was passed by the amplifier. Electrophysiological data was digitized at 40 kHz by a 16-bit A/D converter (USB-6259, National Instruments), low-pass filtered at 10 kHz using the amplifier's internal four-pole Bessel filter, and acquired using pClamp. All electrophysiological data were analyzed using custom Python scripts.

Quantification of Meissner corpuscle and Merkel complex density

Images of serial sections of glabrous pads were sectioned at 25 μm. Images of sections were analyzed in the Fiji distribution of ImageJ (www.fiji.sc). In each section, the number of Meissner corpuscles, visualized by using S100 antibody (Meissner corpuscles formed large, ovoid masses in the dermal papillae compared to Schwann cells surrounding neurons, which are also $S100^+$), and the number of Merkel complexes, visualized by using TROMA-I antibody, were counted. Then, the length of the surface of the skin was traced and measured in ImageJ and multiplied by the thickness of the sections to obtain a volume in mm^3. Density was calculated by dividing the number of corpuscles or Merkel complexes (clusters of Merkel cells) by the volume of skin.

Behavior

Mice used for behavioral experiments were kept on a reversed light-dark cycle and were tested during the dark phase. Male mice of mixed genetic background (C57BL/6J and CD1) were ear notched for identification and genotyped at approximately P21. Females were not used because of potential effects of the estrous cycle on behavior and a preliminary observation that female mice have lower von Frey thresholds than male mice. Ear notching was performed using an ear punch device (Kent Scientific). Mice 6 to 16 weeks of age were subjected to behavioral tests. Mutant animals were compared to control littermates from the same genetic crosses to control for strain or genetic background variability. All behavioral testing was performed by an experimenter blind to genotype.

von Frey paw withdrawal test

For the von Frey paw withdrawal test, a sheet of wire mesh (9217T52, McMaster-Carr) was attached to an acrylic frame and positioned about 14 inches above a table. Mice were placed in 5-inch–by–2-inch clear, acrylic chambers, which were taped on top of the mesh. Before the test date, mice were habituated to the chambers for 1 hour for two consecutive days. On the test date, withdrawal responses were measured. After allowing the mice to habituate for 1 hour, von Frey filaments (North Coast Medical) were applied to the left hindpaw. The experimenter applied the filament onto or within close proximity to the pedal pads of the hindpaw. Care was taken to avoid the hairy skin in the middle of the hindpaw. Starting with the lowest force, each filament was applied five times in a row, followed by a brief break. Then, the filament was applied another five times for a total of 10 applications. The number of paw withdrawals was recorded for each filament weight.

Operant conditioning surgery and water restriction

Before operant conditioning, control *TrkB*$^{flox/flox}$ and mutant *TrkB*^{cKO} mice were implanted with headplates in order to restrain them to the operant conditioning behavior apparatus. Adult (>P60) mice were anaesthetized with isoflurane (1.75 to 2% in O_2). Following removal of the scalp and periosteum, MetaBond dental cement was used to secure a lightweight, titanium headplate to the skull. Five days after surgery, mice began water restriction (40 ml/kg

of initial body weight) until they weighed 85% of their prerestriction weight. Animals were monitored for health conditions every training and testing day using a detailed health assessment adapted for water deprived mice (*39*). Health categories included weight, posture and grooming, activity level, signs of normal eating and waste elimination, and signs of dehydration. The minimum required daily water intake for each animal was 0.8 ml during data collection. Any animal that did not receive this much during the assay was supplemented up to the minimum volume.

Operant conditioning training

Mice were kept on a reversed light-dark cycle, and behavioral sessions were held daily for each animal. Behavioral paradigm control was performed by a custom written program running within the WaveSurfer application (Howard Hughes Medical Institute) in MATLAB. On the first and second days of training, each animal was given about 1 min to freely roam the apparatus platform. Mice were then habituated to the head restraint for 5 to 15 min with ad libitum water from the spout. To minimize the stress and movement of the animal, a three-sided acrylic enclosure was placed over the animal's body. On the third day of training, the right forepaw of each animal was secured over a 3-mm square hole in the apparatus platform, using an elastic strap over the forearm and clear cellophane tape over the back of the forepaw. A fine mechanical stimulator (Model 300 C-I, Aurora Scientific) positioned beneath the platform was used to apply a 75-mN step indentation stimulus to the glabrous skin of the forepaw. This indentation was applied atop a minimal force (<1 mN) used to hold the arm in position. If the animal licked the spout within 2 s of the stimulus onset, it received a 5-μl water reward. Over the next 2 weeks of training, the response period was reduced to 0.5 s, and a no-lick window of 3 s was introduced to discourage spurious licks. In the final phase of training, animals were presented with a range of stimulus amplitudes between 0.5 and 50 mN on 60% of trials, with 20% of all trials serving as no-stimulus (catch) trials, and 20% with a salient 75-mN stimulus to maintain task engagement. Within 3 weeks, animals exhibited stable psychometric functions, with average response rate of 80.5 ± 3.5% and false alarm rate of 17.1 ± 2.5% (mean ± SEM) for the 2 days preceding data collection. Mice displayed some differences in overall sensitivity between von Frey and operant conditioning. Differences in stimulus application—including probe diameter, stimulus waveform, and holding force—likely account for these sensitivity differences.

Operant conditioning data collection and sorting

For all animals, data was collected for 6 days, following a pattern of 2 days of testing blocks followed by 1 day of an easier training paradigm. To exclude data when animals were not engaged in the behavior task, blocks of trials characterized by overall poor performance were omitted from analysis of the detection threshold. We used a d′ statistic to identify these blocks of disengagement in the task, calculated using only trials with salient stimuli levels greater than 40 mN (Eq. 1).

$$d' = Z(\text{hit rate}) - Z(\text{false alarm rate}) \quad (1)$$

For the first and last 14 trials in a session, d′ was calculated over the entire block and the same value was assigned to each of the 14 trials. For all other trials in a session, d′ was calculated over blocks of 30 trials, and the value was assigned to the middle trial in the block (i.e., the d′ value for trial 15 was found using all stimulus trials over 40 mN between trials 1 and 30). Only blocks with at least 15 trials of d′ greater than or equal to 1.5 were used in data analysis. Animals with fewer than 750 viable trials were excluded.

To fit psychometric curves to the aggregate data, all stimulus trials were distributed into bins at 3-mN intervals, with the exception of the final bin with the most salient stimulus of 75 mN. For each animal, the average response rate during trials within each bin was calculated and averaged across all animals in either the control or $TrkB^{cKO}$ group. The MATLAB software package "psignifit 4" was used to fit a beta-binomial logistic sigmoid curve to each group (Eq. 2). Psignifit also output a parameter η that scaled the extra variance introduced within the bins of the data, where values near 1 indicated overdispersed data and values near 0 indicated good binomial dispersion.

$$\Psi(x; m, w, \lambda, \gamma) = \gamma + (1 - \lambda - \gamma)[11 + e^{-2\log(1.05-1)x - mw}] \quad (2)$$

where x is binned stimulus intensity, m is the midpoint of the psychometric function, w is the width of the curve, λ is the upper asymptote, and γ is the lower asymptote. Average psychometric threshold for each group was determined by fitting individual mouse response data to a beta-binomial logistic sigmoid curve using Psignifit. For each mouse, the force value at the midpoint of the psychometric function was determined to be its psychometric threshold. This data was then averaged within group.

Horizontal ladder assay

For the horizontal ladder assay, a ladder consisting of two parallel acrylic sheets as sides and equally spaced aluminum rods was mounted above a table. An angled mirror was positioned below the ladder to increase visibility of foot position. The lighting in the testing room was dim to eliminate visual input as much as possible. Mice were trained on the ladder during two consecutive habituation days. During habituation, mice were placed on one end of the ladder and were allowed to walk across the ladder toward a cage on the opposite end. Trials in which the mice paused for more than 5 s, turned around, or walked on the rim of the ladder rather than the rungs were immediately repeated. This continued until mice successfully crossed the ladder three times, but each mouse was subjected to no more than five total trials per day. On the third day, videos of the mice crossing the ladder were recorded, and the number of slips and time to cross the ladder were measured by an experimenter who was blinded to genotype. The average values of three successful trials were obtained for each animal.

Wire hang assay

For the wire hang assay, a 35-cm aluminum wire (~2.5 mm in diameter) was hung about 14 inches above a table using two aluminum columns attached to a base. Two paper plates were positioned on each end of the wire to prevent mice from moving to the columns. The mouse was placed in the center of the wire, and a cage with bedding was placed below the mouse to provide a cushioned landing area. For each mouse, the time from placement on the wire to falling was measured for three successful trials. Mice that remained on the wire for 90 s or longer were removed from the wire, and their time to fall was counted as 90 s. Additionally, trials in which mice fell instantly after placement on the wire were not counted as successful, and the trial was repeated.

Tactile prepulse inhibition (PPI) assay

Control and $TrkB^{cKO}$ mice were subjected to the tactile inhibition assay, a measure of hairy skin sensitivity, as described (*40*).

Sunflower seed handling measurements

Sunflower seed habituation in home cages: One week before testing animals were habituated to black oil sunflower seeds (Bio-Serv S5137-1 and Wagner's 76025) in their home cage by adding one to two tablespoons of seeds to the floor of the cage for five consecutive days. If animals did not recognize seeds as a food source, a teaspoon of seeds were cracked before adding to the cage floor.

Behavior chamber habituation and handling 2 days before sunflower seed testing was as follows: Animals were habituated to behavior room environment and investigator handling by undergoing tail inking on habituation day 1. To ink the tail, each animal was gently lifted and placed on the cage wire food hopper facing away from the investigator. With the tail firmly held midway from the tail base, a blue permanent soy-ink marker was rolled across the tail, forming parallel lines to indicate identifying ear notch number. Once inked, animals were

gently transferred back into the home cage to await test chamber habituation.

The test chamber was constructed of a black matte acrylic wall and three optically clear walls, 10 inches (*l*)–by–8 inches (*w*)–by–8 inches (*h*) and 0.25 inches thick, centered on a white matte acrylic floor under diffuse warm white light (2700 K). Three digital USB 2.0 CMOS video cameras mounted on camera sliders were positioned on each clear side of the test chambers. One additional overview camera was mounted directly above the test chamber.

Assay and behavioral measurements were as follows: Seeds were withheld from the home cage for 2 days before and during testing to encourage foraging and seed eating in the test chamber. During habituation days 1 and 2, animals were removed from the home cage and placed in an empty test chamber resting on the white matte acrylic floor. Each animal was given two to three seeds while freely exploring the test chamber for 20 min. After the free exploration period, animals were returned to their home cages. On testing day 3, animals were transferred from their home cage and placed in the test chamber and allowed to explore the chamber for 5 min. After acclimation, two to three seeds were placed on the floor of the test chamber and the seed eating activity was recorded. At the completion of the test, animals were removed from the test chamber and returned to their home cage. Chambers were reset and cleaned with unscented soapy water, wiped down with distilled water, and dried. Animals that failed to eat seeds after 20 min were returned to their home cage, and the test was rescheduled. This approach was repeated until each animal fully deshelled and consumed multiple seeds. Seeds that were partially shelled, partially consumed, or discarded were not counted. Behaviors were measured by defined seed deshelling and eating actions: (i) Seed peeling and deshelling—the act of grasping and holding the sunflower seed between the forepaws, clamping the upper and lower incisors into the shell surface, and applying downward force (dip) that pushed the shell away from the head and teeth toward the floor. This action resulted in a systematic peeling of the shell to expose the seed kernel. Animals unable to maintain a firm grip on or fully grasp the shell would typically adapt by touching, tapping, resting and/or bracing the seed against the floor between the forepaws. Animals also "tucked" the shell against their abdomen, holding the shell between the forepaws, clamping their upper and lower incisors onto the shell surface pulling their heads backward away from the shell and forepaws to peel off sections of the shell exposing the seed kernel. (ii) Touch-taps—the act of touching and/or holding and/or bracing the seed shell between the forepaws and the floor during seed peeling. (iii) Dip—the act of holding the seed shell between forepaws, clamping shell between incisors, and applying downward force to peel off sections of shell. (iv) Rotate—the act of or ability to change and/or manipulate shell orientation within the forepaws. (v) Rocking—the act of grasping the shell between the forepaws with shell firmly between incisors, using forepaws to "rock" the shell side to side between the incisors to bite into the shell to peel and expose the seed kernel.

For a table of resources used in the sunflower seed assay, see the supplementary materials.

Gait analysis

A recording setup based on a modified design of LocoMouse (*41*) was built for high-speed, automated analysis of mouse locomotion behavior. The dimensions of the transparent corridor used were 64.5 cm (*l*) by 4 cm (*w*) by 6 cm (*h*). Dark enclosures were situated on both ends of the corridor, and the animal freely moved between the two ends. Acquisition were triggered by infrared sensors during each epoch of corridor crossing. Two mirrors angled at 48° downward flanked the longitudinal axis of the corridor, which projected the side views of the animal to a camera below. A single high-speed camera (Bonito CL-400B/C 2320 × 700 pixels, Allied Vision, Exton, PA) captured simultaneous videos of the bottom view and two side views of the animal at 200 frames per second. Videos of the animal were compressed and analyzed offline using a convolutional neural network for postural tracking written in PyTorch and MATLAB (Mathworks, Natick, MA). Briefly, an hourglass network based on (*42*) was trained on 1000 expert annotated frames to simultaneously recognize the following body parts in the bottom and side views: nose, start and end of the tail, forepaws, and hind paws. The positions of body parts from all three perspectives were combined to generate a three-dimensional postural time series. A hidden Markov model was used to infer the start and the end of each gait cycle based on paw velocity. The following parameters were measured for each gait cycle: velocity, height, tail elevation, hind paw width, cadence, stride length, paw width, stand duration, and vertical and horizontal tail oscillation. A linear mixed effect model of the form Y ~ Intercept + Genotype + Sex + Genotype:Sex+ (1|Name) and ANOVA was used for statistical testing. In the case where velocity is considered as an independent variable, a model of the form Y ~ Intercept + Velocity + Genotype + Sex + Genotype:Sex + Velocity:Genotype + Velocity:Sex + (1|Name) was used instead.

Electron microscopy

Forepaw toe pads (ages P21 to P39) were excised and immersed in a glutaraldehyde-formaldehyde fixative for 1 hour at room temperature, further dissected, and subsequently fixed overnight at 4°C. Sample preparation was done as previously described (*28*). Ultrathin sections were cut at 40 to 60 nm and imaged using a JEOL 1200EX transmission electron microscope at 80-kV accelerating voltage. Images were cropped and adjusted with normalization to enhance contrast using Fiji/ImageJ. For corpuscles that were larger than the imaging field, montage of images was generated using TrakEM2 (*43*). Only sections close to the center of Meissner corpuscles were included for quantification. Axons that did not have apparent lamellar wrappings were not included. Two researchers, not involved in EM data collection and blind to genotypes, independently counted the number of lamellar wrappings around designated axons. The counts from both researchers were averaged for final analysis. Differences in lamellar wrapping counts per axon between these two researchers ranged from 0 to 5 wrappings.

Modeling

To find the mutual information between point stimulus location and neural population response, we assumed K possible point stimulus locations, denoted s, where each of them is equally likely to occur, that is, $p(s) = 1/K$ for all s. Furthermore, we assumed $M(\leq K)$ distinct neural population responses, denoted r, where response $r = m$ corresponds to n_m different point stimulus locations. Some locations might not have tactile receptors, such that $L = \sum_{m=1}^{M} n_m \leq K$. Overall, the set $\{n_1, \ldots, n_M\}$ determines the efficiency with which neural population responses encode tactile location.

To find the mutual information between point stimulus location s and neural population response r, we use $\mathrm{MI}(s;r) = \mathrm{H}(s) - \mathrm{H}(s|r)$, where $\mathrm{H}(s)$ is the entropy of stimulus locations, and $\mathrm{H}(s|r)$ is the conditional entropy, conditioned on a particular population response. Because of the uniform stimulus location distribution, we have $\mathrm{H}(s) = \log_2 K$. To find the conditional entropy, we observe that $p(s|r = m) = 1/n_m$ for all locations s that yield population response $r = m$, and $p(s|r = m) = 0$ otherwise. This results in

$$\mathrm{H}(s|r) = -\sum_s p(s) \sum_{m=1}^{M} p(s|r = m) \log_2 p(s|r = m) = -\sum_{m=1}^{M} \frac{n_m}{K} \log_2 \frac{1}{n_m}$$

where we have used the fact that each stimulus location corresponds to at most one population response. Overall, this yields the mutual information

$$\mathrm{MI}(s;r) = \log_2 K + \sum_{m=1}^{M} \frac{n_m}{K} \log_2 \frac{1}{n_m}$$

In this expression, n_m/K is the overall fraction of stimulus locations that evoke population response $r = m$, and $1/n_m$ is the probability of each stimulus location given this response.

To find the set $\{n_1, ..., n_M\}$ that maximizes the mutual information, we define the column vector $\vec{n} = (n_1, ..., n_M)^T$ and write the mutual information in vector form, $\mathrm{MI}(s;r) = \log_2 K + K^{-1}\vec{n}^T \log_2 1/\vec{n}$, where the last division is element-wise. We can then perform a constraint optimization on $\vec{n}$ with constraint $\vec{1}^T\vec{n} = L$, resulting in $n_m^* = L/M$ for all m. Thus, optimal coding (under the above assumptions) distributes the population responses evenly across stimulus locations. Furthermore, the mutual information increases with L, such that it is maximized if all stimulus locations yield a neural response (i.e., there are no gaps in the receptive fields), when $L^* = K$. Under these circumstances, the mutual information becomes $\mathrm{MI}(s;r) = \log_2 M$, providing an upper bound.

For nonoverlapped, square receptive fields, the number of distinct population responses M equals the number of neurons N. If these receptive fields tile the whole stimulus space evenly, then the associated mutual information is $\mathrm{MI}(s;r) = \log_2 N$. As the number of discriminable spaces is $2^{MI(s;r)} = N$, the number of efficiently discriminable spaces per neuron is $2^{MI(s;r)}/N = 1$.

For overlapped, square receptive fields, each stimulus location within the receptive field of two distinct neurons results in a distinct neural population response. Two nonoverlapped layers with $N/2$ equally sized square receptive fields that are offset against each other result in $4(N/2) = 2N$ distinct population responses (ignoring boundary effects, which become negligible for large N). These responses evenly tile the stimulus space if the offset is maximized, in which case $\mathrm{MI}(s;r) = \log_2 2N$. This results the number of efficiently discriminable space per neuron to be given by $2^{MI(s;r)}/N = 2$.

The two-point discrimination capability of square receptive fields was modeled by imagining a point stimulus presented to the center of a discriminable space and a circle of points presented at some distance d from the central point. The central point and the shifted point were presented asynchronously. For a given d, the fraction of the circle's circumference that extended outside the central discriminable space into a new discriminable space was considered the percent correct (i.e., the percentage of points unambiguously distinguishable from the central point). For nonoverlapped squares with side length s and with a central point presented within a single square

$$\%\ \text{correct} = \begin{cases} 0, & d \le \frac{s}{2} \\ 100\frac{4}{\pi}\cos^{-1}\frac{s}{2d}, & \frac{s}{2} < d \le \frac{s}{\sqrt{2}} \\ 100, & d < \frac{s}{\sqrt{2}} \end{cases}$$

For overlapped squares with side length $\sqrt{2}s$ and with a central point presented within a discriminable space:

$$\%\ \text{correct} = \begin{cases} 0, & d \le \frac{s}{2\sqrt{2}} \\ 100\frac{4}{\pi}\cos^{-1}\frac{s}{2\sqrt{2}d}, & \frac{s}{2\sqrt{2}} < d \le \frac{s}{2} \\ 100, & d < \frac{s}{2} \end{cases}$$

To model irregular and variable receptive fields, images of overlapped and nonoverlapped Voronoi tessellations were generated in MATLAB using the Voronoi function. The image set consisted of 4000 tessellations seeded with 84 to 101 random points in a square window. Number of neurons, N, for a single tessellation was the number of seeded points forming Voronoi cells that were within the window. Pairs of tessellations were overlapped to create 2000 overlapped mosaics with number of neurons equal to the combined number of neurons in the overlaid pairs. Mutual information between neural responses r and stimuli s was defined as above:

$$\begin{aligned} MI(s;r) &= H(s) - H(s|r) \\ &= \log_2 K + \sum\nolimits_{m=1}^{M} \frac{n_m}{K} \log_2 \frac{1}{n_m} \end{aligned}$$

where K is the total number of pixels and n_m are the number of pixels per discriminable space. Mutual information was normalized by the number of neurons ($2^{MI}/N$). The difference between overlapped and nonoverlapped Voronoi tessellations was determined to be statistically significant using 5000 bootstrap samples, each consisting of a randomly chosen pair from the 2000 overlapped tessellations and 2000 nonoverlapped tessellations.

Code availability

The computer code that supports the findings of this study is available from the corresponding author upon reasonable request.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES AND NOTES

1. R. Wagner, G. Meissner, Über das Vorhandensein bisher unbekannter eigenthümlicher Tastkörperchen (Corpuscula tactus) in den Gefühlswarzchen der menschlichen Haut and über die Endausbreitung sensitiver Nerven. *Nachricten von der Georg-August-Universität und der Königl Gesellschaft der Wissenschaften zu Göttingen* **2**, 17–30 (1852).
2. T. González-Martínez *et al*., Absence of Meissner corpuscles in the digital pads of mice lacking functional TrkB. *Brain Res.* **1002**, 120–128 (2004). doi: 10.1016/j.brainres.2004.01.003; pmid: 14988041
3. T. González-Martínez *et al*., BDNF, but not NT-4, is necessary for normal development of Meissner corpuscles. *Neurosci. Lett.* **377**, 12–15 (2005). doi: 10.1016/j.neulet.2004.11.078; pmid: 15722178
4. P. Perez-Pinera *et al*., Characterization of sensory deficits in *TrkB* knockout mice. *Neurosci. Lett.* **433**, 43–47 (2008). doi: 10.1016/j.neulet.2007.12.035; pmid: 18248898
5. S. da Silva *et al*., Proper formation of whisker barrelettes requires periphery-derived Smad4-dependent TGF-β signaling. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3395–3400 (2011). doi: 10.1073/pnas.1014411108; pmid: 21300867
6. Y. Liu *et al*., Sexually dimorphic BDNF signaling directs sensory innervation of the mammary gland. *Science* **338**, 1357–1360 (2012). doi: 10.1126/science.1228258; pmid: 23224557
7. L. Bai *et al*., Genetic identification of an expansive mechanoreceptor sensitive to skin stroking. *Cell* **163**, 1783–1795 (2015). doi: 10.1016/j.cell.2015.11.060; pmid: 26687362
8. C. Ma, D. F. Donnelly, R. H. LaMotte, In vivo visualization and functional characterization of primary somatic neurons. *J. Neurosci. Methods* **191**, 60–65 (2010). doi: 10.1016/j.jneumeth.2010.06.010; pmid: 20558205
9. W. Luo, H. Enomoto, F. L. Rice, J. Milbrandt, D. D. Ginty, Molecular identification of rapidly adapting mechanoreceptors and their developmental dependence on ret signaling. *Neuron* **64**, 841–856 (2009). doi: 10.1016/j.neuron.2009.11.003; pmid: 20064391
10. S. M. Maricich, K. M. Morrison, E. L. Mathes, B. M. Brewer, Rodents rely on Merkel cells for texture discrimination tasks. *J. Neurosci.* **32**, 3296–3300 (2012). doi: 10.1523/JNEUROSCI.5307-11.2012; pmid: 22399751
11. S. M. Maricich *et al*., Merkel cells are essential for light-touch responses. *Science* **324**, 1580–1582 (2009). doi: 10.1126/science.1172890; pmid: 19541997
12. K. M. Morrison, G. R. Miesegaes, E. A. Lumpkin, S. M. Maricich, Mammalian Merkel cells are descended from the epidermal lineage. *Dev. Biol.* **336**, 76–83 (2009). doi: 10.1016/j.ydbio.2009.09.032; pmid: 19782676
13. C. N. Perdigoto, E. S. Bardot, V. J. Valdes, F. J. Santoriello, E. Ezhkova, Embryonic maturation of epidermal Merkel cells is controlled by a redundant transcription factor network. *Development* **141**, 4690–4696 (2014). doi: 10.1242/dev.112169; pmid: 25468937
14. A. Ramirez *et al*., A keratin K5Cre transgenic line appropriate for tissue-specific or generalized Cre-mediated recombination. *Genesis* **39**, 52–57 (2004). doi: 10.1002/gene.20025; pmid: 15124227
15. N. F. Shroyer *et al*., Intestine-specific ablation of *mouse atonal homolog 1* (*Math1*) reveals a role in cellular homeostasis. *Gastroenterology* **132**, 2478–2488 (2007). doi: 10.1053/j.gastro.2007.03.047; pmid: 17570220
16. A. Van Keymeulen *et al*., Epidermal progenitors give rise to Merkel cells during embryonic development and adult homeostasis. *J. Cell Biol.* **187**, 91–100 (2009). doi: 10.1083/jcb.200907080; pmid: 19786578
17. L. Madisen *et al*., A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010). doi: 10.1038/nn.2467; pmid: 20023653
18. M. Rutlin *et al*., The cellular and molecular basis of direction selectivity of Aδ-LTMRs. *Cell* **159**, 1640–1651 (2014). doi: 10.1016/j.cell.2014.11.038; pmid: 25525881
19. T. Uesaka, M. Nagashimada, S. Yonemura, H. Enomoto, Diminished *Ret* expression compromises neuronal survival in the colon and causes intestinal aganglionosis in mice. *J. Clin. Invest.* **118**, 1890–1898 (2008). doi: 10.1172/JCI34425; pmid: 18414682
20. L. Li *et al*., The functional organization of cutaneous low-threshold mechanosensory neurons. *Cell* **147**, 1615–1627 (2011). doi: 10.1016/j.cell.2011.11.027; pmid: 22196735
21. V. E. Abraira, D. D. Ginty, The sensory neurons of touch. *Neuron* **79**, 618–639 (2013). doi: 10.1016/j.neuron.2013.07.051; pmid: 23972592
22. A. Iggo, H. Ogawa, Correlative physiological and morphological studies of rapidly adapting mechanoreceptors in cat's glabrous skin. *J. Physiol.* **266**, 275–296 (1977). doi: 10.1113/jphysiol.1977.sp011768; pmid: 853451
23. W. D. Willis, R. E. Coggeshall, *Sensory Mechanisms of the Spinal Cord* (Kluwer Academic/Plenum Publishers, ed. 3, 2004), vol. 1.
24. K. O. Johnson, The roles and functions of cutaneous mechanoreceptors. *Curr. Opin. Neurobiol.* **11**, 455–461 (2001). doi: 10.1016/S0959-4388(00)00234-8; pmid: 11502392
25. A. G. Brown, R. E. Fyffe, R. Noble, Projections from Pacinian corpuscles and rapidly adapting mechanoreceptors of glabrous skin to the cat's spinal cord. *J. Physiol.* **307**, 385–400 (1980). doi: 10.1113/jphysiol.1980.sp013441; pmid: 7205669
26. L. M. Pubols, B. H. Pubols Jr.., Modality composition and functional characteristics of dorsal column mechanoreceptive afferent fibers innervating the raccoon's forepaw. *J. Neurophysiol.* **36**, 1023–1037 (1973). doi: 10.1152/jn.1973.36.6.1023; pmid: 4761717
27. P. Shortland, C. J. Woolf, Morphology and somatotopy of the central arborizations of rapidly adapting glabrous skin afferents in the rat lumbar spinal cord. *J. Comp. Neurol.* **329**, 491–511 (1993). doi: 10.1002/cne.903290406; pmid: 8454737

28. Q. Zhang, W. A. Lee, D. L. Paul, D. D. Ginty, Multiplexed peroxidase-based electron microscopy labeling enables simultaneous visualization of multiple cell types. *Nat. Neurosci.* **22**, 828–839 (2019). doi: 10.1038/s41593-019-0358-7; pmid: 30886406
29. W. R. Loewenstein, M. Mendelson, Components of receptor adaptation in a Pacinian corpuscle. *J. Physiol.* **177**, 377–397 (1965). doi: 10.1113/jphysiol.1965.sp007598; pmid: 14321486
30. B. M. Hooks, J. Y. Lin, C. Guo, K. Svoboda, Dual-channel circuit mapping reveals sensorimotor convergence in the primary motor cortex. *J. Neurosci.* **35**, 4418–4426 (2015). doi: 10.1523/JNEUROSCI.3741-14.2015; pmid: 25762684
31. T. C. Badea *et al*., New mouse lines for the analysis of neuronal morphology using CreER(T)/loxP-directed sparse labeling. *PLOS ONE* **4**, e7859 (2009). doi: 10.1371/journal.pone.0007859; pmid: 19924248
32. L. Madisen *et al*., Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015). doi: 10.1016/j.neuron.2015.02.022; pmid: 25741722
33. H. Wu, J. Williams, J. Nathans, Morphologic diversity of cutaneous sensory afferents revealed by genetically directed sparse labeling. *eLife* **1**, e00181 (2012). doi: 10.7554/eLife.00181; pmid: 23256042
34. J. A. Gorski, S. R. Zeiler, S. Tamowski, K. R. Jones, Brain-derived neurotrophic factor is required for the maintenance of cortical dendrites. *J. Neurosci.* **23**, 6856–6865 (2003). doi: 10.1523/JNEUROSCI.23-17-06856.2003; pmid: 12890780
35. P. S. Danielian, D. Muccino, D. H. Rowitch, S. K. Michael, A. P. McMahon, Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Curr. Biol.* **8**, 1323–1326 (1998). doi: 10.1016/S0960-9822(07)00562-3; pmid: 9843687
36. M. Jaegle *et al*., The POU proteins Brn-2 and Oct-6 share important functions in Schwann cell development. *Genes Dev.* **17**, 1380–1391 (2003). doi: 10.1101/gad.258203; pmid: 12782656
37. T. C. Badea, H. Cahill, J. Ecker, S. Hattar, J. Nathans, Distinct roles of transcription factors Brn3a and Brn3b in controlling the development, morphology, and function of retinal ganglion cells. *Neuron* **61**, 852–864 (2009). doi: 10.1016/j.neuron.2009.01.020; pmid: 19323995
38. S. Gong *et al*., A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003). doi: 10.1038/nature02033; pmid: 14586460
39. Z. V. Guo *et al*., Flow of cortical activity underlying a tactile decision in mice. *Neuron* **81**, 179–194 (2014). doi: 10.1016/j.neuron.2013.10.020; pmid: 24361077
40. L. L. Orefice *et al*., Peripheral mechanosensory neuron dysfunction underlies tactile and behavioral deficits in mouse models of ASDs. *Cell* **166**, 299–313 (2016). doi: 10.1016/j.cell.2016.05.033; pmid: 27293187
41. A. S. Machado, D. M. Darmohray, J. Fayad, H. G. Marques, M. R. Carey, A quantitative framework for whole-body coordination reveals specific deficits in freely walking ataxic mice. *eLife* **4**, e07892 (2015). doi: 10.7554/eLife.07892; pmid: 26433022
42. A. Newell, K. Yang, J. Deng, Stacked hourglass networks for human pose estimation. arXiv:1603.06937v2 [cs.CV] (26 July 2016).
43. A. Cardona *et al*., TrakEM2 software for neural circuit reconstruction. *PLOS ONE* **7**, e38011 (2012). doi: 10.1371/journal.pone.0038011; pmid: 22723842

ACKNOWLEDGMENTS

We thank E. Kuehn for AAV virus characterization, W.-C. Lee for providing access to ultramicrotomy equipment, and R. LaMotte and J. Woodbury for assistance with the in vivo physiological recording preparation. We thank M. Rutlin for preliminary analysis of behavioral deficits in *TrkB* conditional mutant mice, J. Barowski for help with figures, D. Paul for help with EM quantification, and T. Monteiro for technical assistance. We thank O. Mazor and P. Gorelik of the HMS Research Instrumentation Core Facility for consultation on operant conditioning design and MATLAB code and J. Huang for discussing unpublished sunflower seed handling measurements. We thank M. Pecot, S. Ashrafi, N. Sharma, and G. Fishell for helpful comments on the manuscript. **Funding:** N.L.N., M.W.S., Q.Z., and C.G. were Stuart H.Q. and Victoria Quan fellows at Harvard Medical School. This work was supported by NIH grants NS97344 (D.D.G.), NS105324 (A.J.E.), NS101843 (M.W.S), MH115554 (J.D.), K99 NS101057 (L.L.O.), and NS089521 (C.D.H.); NSF GRFP 2014177995 (M.W.S.); a scholar award by the James S. McDonnell Foundation (J.D.); and the Edward R. and Anne G. Lefler Center for Neurodegenerative Disorders (D.D.G.). D.D.G. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** N.L.N., Y.L., A.J.E., and D.D.G. conceived the project; N.L.N. and Y.L. performed light microscopy experiments with assistance from A.A.; N.L.N., A.J.E., and B.P.L. did in vivo electrophysiological recordings with guidance from C.D.H.; M.W.S., M.I., J.R., and S.J.K. performed the operant conditioning behavioral experiments; L.L.O. performed the PPI experiments; C.G. and M.M.D. performed gait analysis with guidance from W.R.; N.L.N., A.A., and M.M.D. performed all other behavioral experiments; N.L.N. and J.D. performed the modeling; A.H., Q.Z., and S.J.C. did the EM analysis; N.L.N. and D.D.G. wrote the first draft of the paper, with sections subsequently provided by A.J.E.; and all authors contributed to the final manuscript draft. **Competing interests:** The authors have no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. The mouse strains generated for and used in this study are available from the corresponding author under a material agreement with Harvard.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/eabb2751/suppl/DC1
Supplementary Text
Figs. S1 to S10
MDAR Reproducibility Checklist
Movies S1 and S2

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

12 February 2020; accepted 28 April 2020
10.1126/science.abb2751

RESEARCH ARTICLE

CORONAVIRUS

Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease

Wenhao Dai[1,2]*, Bing Zhang[3]*, Xia-Ming Jiang[4]*, Haixia Su[1,5]*, Jian Li[1,6], Yao Zhao[3], Xiong Xie[1,5], Zhenming Jin[3], Jingjing Peng[1,5], Fengjiang Liu[3], Chunpu Li[1,5], You Li[7], Fang Bai[3], Haofeng Wang[3], Xi Cheng[1], Xiaobo Cen[7], Shulei Hu[1], Xiuna Yang[3], Jiang Wang[1,5], Xiang Liu[8], Gengfu Xiao[4], Hualiang Jiang[1,2,3,5], Zihe Rao[3], Lei-Ke Zhang[4]†, Yechun Xu[1,5]†, Haitao Yang[3]†, Hong Liu[1,2,5,6]†

SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) is the etiological agent responsible for the global COVID-19 (coronavirus disease 2019) outbreak. The main protease of SARS-CoV-2, M^{pro}, is a key enzyme that plays a pivotal role in mediating viral replication and transcription. We designed and synthesized two lead compounds (11a and 11b) targeting M^{pro}. Both exhibited excellent inhibitory activity and potent anti–SARS-CoV-2 infection activity. The x-ray crystal structures of SARS-CoV-2 M^{pro} in complex with 11a or 11b, both determined at a resolution of 1.5 angstroms, showed that the aldehyde groups of 11a and 11b are covalently bound to cysteine 145 of M^{pro}. Both compounds showed good pharmacokinetic properties in vivo, and 11a also exhibited low toxicity, which suggests that these compounds are promising drug candidates.

In late December 2019, a cluster of pneumonia cases caused by a novel coronavirus was reported in Wuhan, China (*1*–*3*). Genomic sequencing showed that this pathogenic coronavirus is 96% identical to a bat coronavirus and shares 79.6% sequence identity to SARS-CoV (*4*–*6*). This novel coronavirus was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses, and the pneumonia was designated as coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO) on 11 February 2020 (*7*). The epidemic spread rapidly to more than 212 countries and was announced as a global health emergency by WHO (*8*). No clinically effective vaccines or specific antiviral drugs are currently available for the prevention and treatment of COVID-19 infections. The combination of α-interferon and the anti-HIV drugs lopinavir/ritonavir (Kaletra) has been used, but the curative effect remains very limited and there can be toxic side effects (*9*). Remdesivir, a broad-spectrum antiviral drug developed by Gilead Sciences Inc., is also being explored for the treatment of COVID-19, but more data are needed to prove its efficacy (*10*–*12*). Specific anti–SARS-CoV-2 drugs offering efficacy and safety are urgently needed.

A maximum likelihood tree based on the genomic sequence showed that the virus falls within the subgenus *Sarbecovirus* of the genus *Betacoronavirus* (*6*). Coronaviruses are enveloped, positive-sense, single-stranded RNA viruses. The genomic RNA of coronaviruses is ~30,000 nucleotides in length with a 5′-cap structure and a 3′-poly(A) tail, and contains at least six open reading frames (ORFs) (*13*, *14*). The first ORF (ORF 1a/b), about two-thirds of the genome length, directly translates two polyproteins, pp1a and pp1ab, so named because there is an a-1 frameshift between ORF1a and ORF1b. These polyproteins are processed by a main protease, M^{pro} [also known as the 3C-like protease ($3CL^{pro}$)], and by one or two papain-like proteases, into 16 nonstructural proteins (NSPs). These NSPs engage in the production of subgenomic RNAs that encode four main

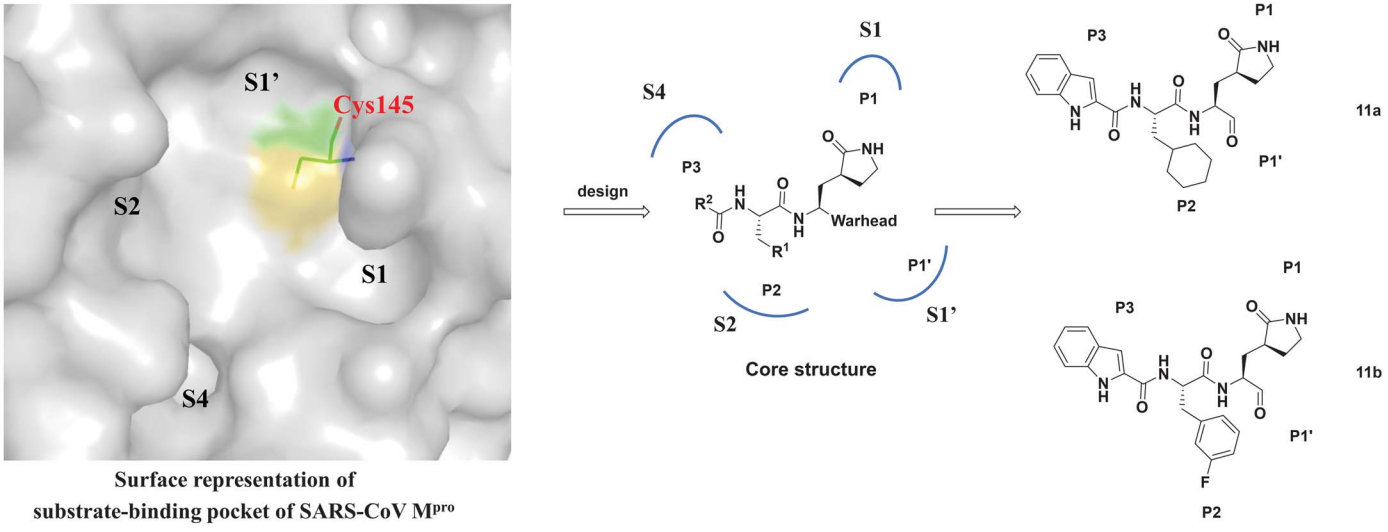


Fig. 1. Design strategy of SARS-CoV-2 main protease (M^{pro}) inhibitors and the chemical structures of 11a and 11b.

[1]State Key Laboratory of Drug Research, CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China. [2]School of Pharmacy, China Pharmaceutical University, Nanjing 210009, Jiangsu, China. [3]Shanghai Institute for Advanced Immunochemical Studies and School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China. [4]State Key Laboratory of Virology, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan 430071, China. [5]University of Chinese Academy of Sciences, Beijing 100049, China. [6]College of Pharmacy, Nanjing University of Chinese Medicine, Qixia District, Nanjing 210023, China. [7]National Chengdu Center for Safety Evaluation of Drugs, Westchina Hospital of Sichuan University, High-Tech Development Zone, Chengdu, Sichuan 610041, China. [8]State Key Laboratory of Medicinal Chemical Biology, Frontiers Science Center for Cell Response, College of Life Sciences, College of Pharmacy, Nankai University, Tianjin 300353, China.
*These authors contributed equally to this work.
†Corresponding author. Email: hliu@simm.ac.cn (H.L.); yanght@shanghaitech.edu.cn (H.Y.); zhangleike@wh.iov.cn (L.-K.Z.); ycxu@simm.ac.cn (Y.X.)

structural proteins [envelope (E), membrane (M), spike (S), and nucleocapsid (N) proteins] and other accessory proteins (*15*, *16*). Therefore, these proteases, especially M^{pro}, play a vital role in the life cycle of coronaviruses.

M^{pro} is a three-domain (domains I to III) cysteine protease involved in most maturation cleavage events within the precursor polyprotein (*17*–*19*). Active M^{pro} is a homodimer containing two protomers. The coronavirus M^{pro} features a noncanonical Cys-His dyad located in the cleft between domains I and II (*17*–*19*). M^{pro} is conserved among coronaviruses, and several common features are shared among the substrates of M^{pro} in different coronaviruses. The amino acids in substrates from the N terminus to the C terminus are numbered as follows: -P4-P3-P2-P1↓P1′-P2′-P3′-, with the cleavage site between P1 and P1′. In particular, a Gln is almost always required in the P1 position of the substrates. Because M^{pro} has no human homolog, it is an ideal antiviral target (*20*–*22*).

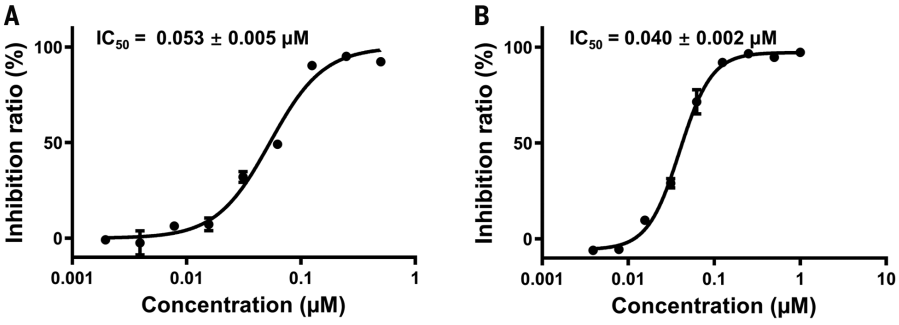


Fig. 2. Inhibitory activity profiles of compounds 11a (A) and 11b (B) against SARS-CoV-2 M^{pro}.

Design and synthesis of 11a and 11b

The active sites of M^{pro} are highly conserved among all coronavirus M^{pro}s and are usually composed of four sites: S1′, S1, S2, and S4 (*22*). We were able to design and synthesize inhibitors targeting SARS-CoV-2 M^{pro} by analyzing the substrate-binding pocket of SARS-CoV M^{pro} (PDB ID 2H2Z) (Fig. 1). The thiol of a cysteine residue in the S1′ site anchors inhibitors by a covalent linkage that is important for the inhibitors to maintain antiviral activity. In our design of new inhibitors, an aldehyde was selected as a new warhead in P1 in order to form

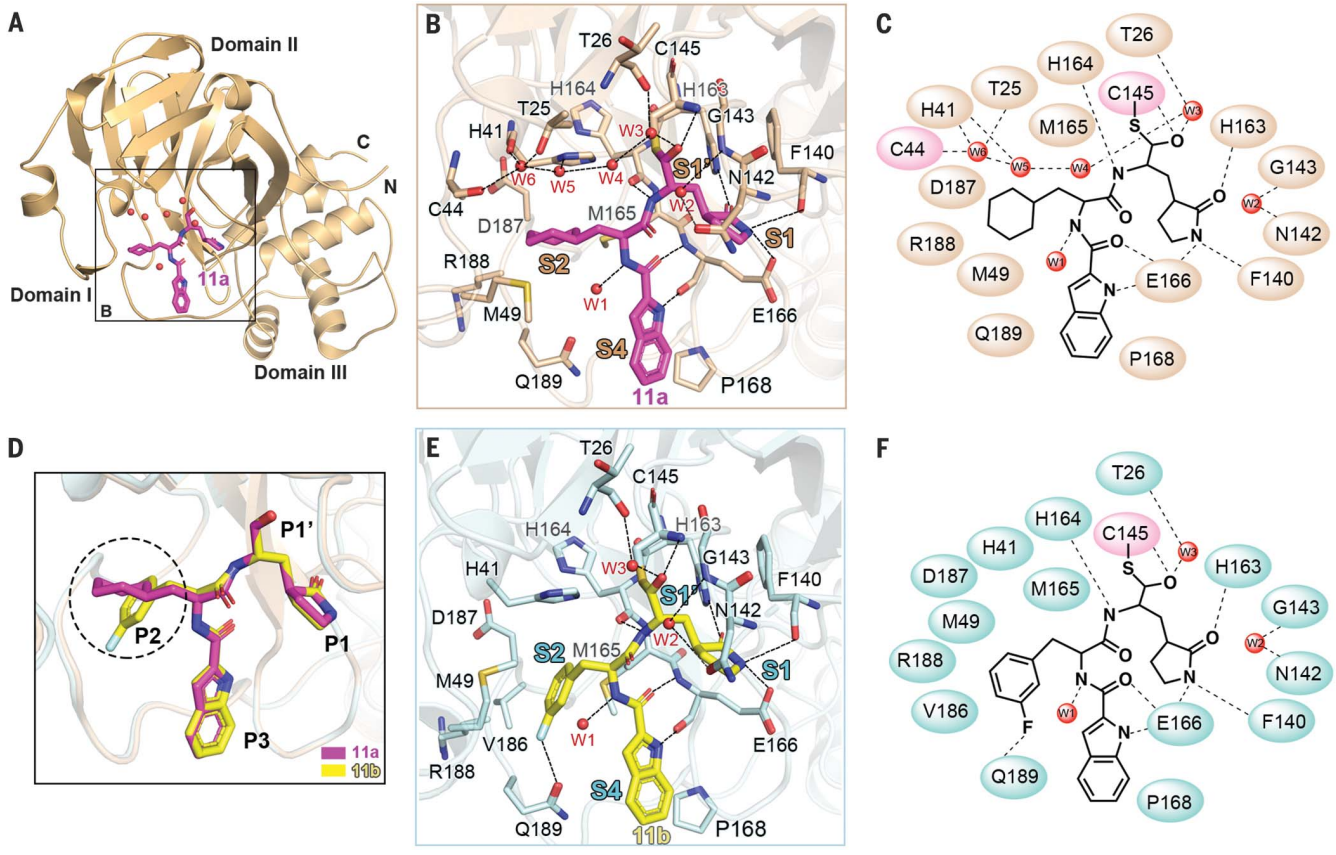


Fig. 3. M^{pro}-inhibitor binding modes for 11a and 11b. (**A**) Cartoon representation of the crystal structure of SARS-CoV-2 M^{pro} in complex with **11a**. Compound **11a** is shown as magenta sticks; water molecules are shown as red spheres. (**B**) Close-up view of the **11a** binding pocket. Four subsites, S1′, S1, S2, and S4, are labeled. The residues involved in inhibitor binding are shown as wheat sticks. **11a** and water molecules are shown as magenta sticks and red spheres, respectively. Hydrogen bonds are indicated as dashed lines. (**C**) Schematic diagram of SARS-CoV-2 M^{pro}–**11a** interactions shown in (B). (**D**) Comparison of the binding modes of **11a** and **11b** in SARS-CoV-2 M^{pro}. The major difference between **11a** and **11b** is marked with a dashed circle; **11a** and **11b** are shown as magenta and yellow sticks, respectively. (**E**) Close-up view of the **11b** binding pocket. Hydrogen bonds are indicated as dashed lines. (**F**) Schematic diagram of SARS-CoV-2 M^{pro}–**11b** interactions shown in (E). Amino acid abbreviations: 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; Y, Tyr.

a covalent bond with cysteine. The reported SARS-CoV M^{pro} inhibitors often have an (*S*)-γ-lactam ring that occupies the S1 site of M^{pro}, and this ring was expected to be a good choice in P1 (*23*). Furthermore, the S2 site of coronavirus M^{pro} is usually large enough to accommodate the larger P2 fragment. To test the importance of different ring systems, we introduced a cyclohexyl or 3-fluorophenyl into P2, with the fluorine expected to enhance activity. An indole group was introduced into P3 to form new hydrogen bonds with S4 and improve drug-like properties.

The synthetic route and chemical structures of the compounds (**11a** and **11b**) are shown in scheme S1. The starting material (*N*-Boc-L-glutamic acid dimethyl ester **1**) was obtained from commercial suppliers and used without further purification to synthesize the key intermediate ***3*** according to the literature (*24*). The intermediates **6a** and **6b** were synthesized from **4** and acids **5a** and **5b**. Removal of the *t*-butoxycarbonyl group from **6a** and **6b** yielded **7a** and **7b**. Coupling **7a** and **7b** with the acid **8** yielded the esters **9a** and **9b**. The peptidomimetic aldehydes **11a** and **11b** were approached through a two-step route in which the ester derivatives **9** were first reduced with $NaBH_4$ to generate the primary alcohols **10a** and **10b**, which were subsequently oxidized into aldehydes **11a** and **11b** with Dess-Martin periodinane (DMP).

11a and 11b are potent inhibitors of M^{pro}

Recombinant SARS-CoV-2 M^{pro} was expressed and purified from *Escherichia coli* (*18*, *25*). A fluorescently labeled substrate—MCA-AVLQ↓SGFR-Lys(Dnp)-Lys-NH_2, derived from the N-terminal autocleavage sequence from the viral protease—was designed and synthesized for the enzymatic assay.

Both **11a** and **11b** exhibited high SARS-CoV-2 M^{pro} inhibition activity, which reached 100% for **11a** and 96% for **11b** at 1 μM, respectively. We used a fluorescence resonance energy transfer (FRET)–based cleavage assay to determine the median inhibitory concentration (IC_{50}) values. The results revealed excellent inhibitory potency, with IC_{50} values of 0.053 ± 0.005 μM and 0.040 ± 0.002 μM for **11a** and **11b**, respectively (Fig. 2).

Structures of SARS-CoV-2 M^{pro} in complex with 11a and 11b

To elucidate the mechanism of inhibition of SARS-CoV-2 M^{pro} by **11a**, we determined the crystal structure of this complex at 1.5 Å resolution (table S1). The crystal of M^{pro}-**11a** belongs to the space group *C*2, and an asymmetric unit contains only one molecule (table S1). Two molecules (designated protomer A and protomer B) associate into a homodimer around a crystallographic two-fold symmetry axis (fig. S2). The structure of each protomer contains three domains with the substrate-binding site located in the cleft between domains I and II. At the active site of SARS-CoV-2 M^{pro}, Cys^{145} and His^{41} (Cys-His) form a catalytic dyad (fig. S2).

The electron density map clearly shows compound **11a** in the substrate-binding pocket of SARS-CoV-2 M^{pro} in an extended conformation (Fig. 3A and fig. S3, A and B). Details

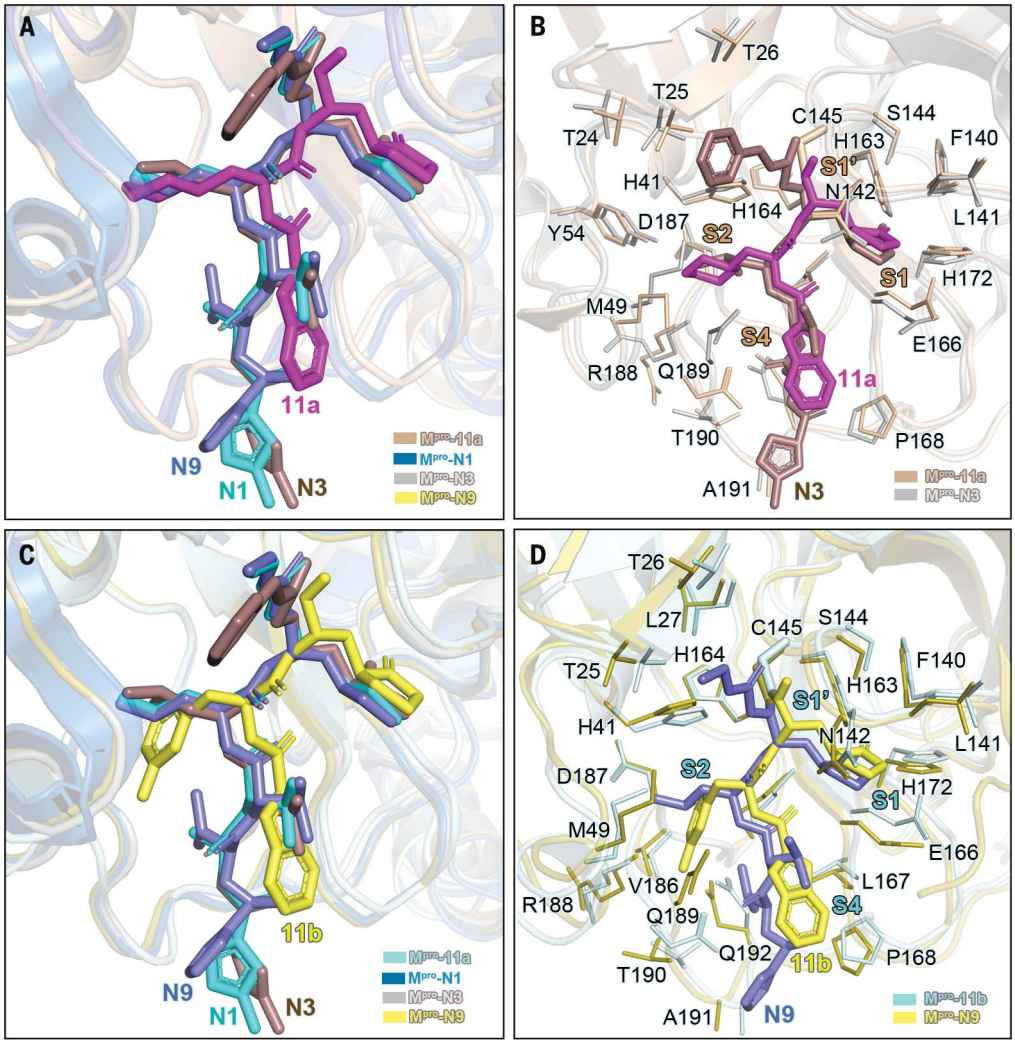


Fig. 4. Comparison of the inhibitor binding modes in SARS-CoV M^{pro} and SARS-CoV-2 M^{pro}. (**A**) Comparison of binding modes of **11a** in SARS-CoV-2 M^{pro} with those of **N1**, **N3**, and **N9** in SARS-CoV M^{pro}. SARS-CoV-2 M^{pro}–**11a** (wheat; PDB code 6LZE), SARS-CoV M^{pro}–**N1** (sky blue; PDB code 1WOF), SARS-CoV M^{pro}–**N3** (gray; PDB code 2AMQ), and SARS-CoV M^{pro}–**N9** (olive; PDB code 2AMD) are shown as cartoons. **11a**, **N1**, **N3**, and **N9** are shown in magenta, cyan, dirty violet, and salt, respectively. (**B**) Comparison of the **11a** and **N3** binding pockets. Residues in the M^{pro}-**11a** and M^{pro}-**N3** structures are colored in wheat and gray, respectively. **11a** and **N3** are shown as sticks colored in magenta and dirty violet, respectively. (**C**) Comparison of binding modes of **11b** in SARS-CoV-2 M^{pro} with those of **N1**, **N3**, and **N9** in SARS-CoV M^{pro}. SARS-CoV-2 M^{pro}–**11b** (PDB code 6M0K) is shown as a pale cyan cartoon. **11b**, **N1**, **N3**, and **N9** are shown in yellow, cyan, dirty violet, and salt, respectively. (**D**) Comparison of the **11b** and **N9** binding pockets. Residues in the M^{pro}-**11b** and M^{pro}-**N9** structures are colored in pale cyan and olive, respectively. **11b** and **N9** are shown as sticks colored in yellow and salt, respectively.

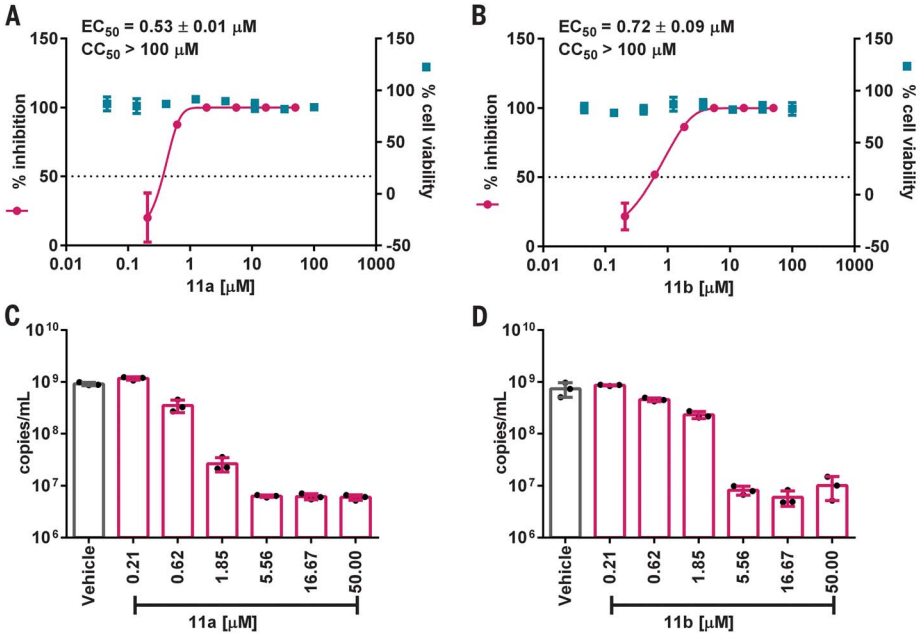


Fig. 5. In vitro inhibition of viral main protease inhibitors against SARS-CoV-2. (**A** and **B**) Vero E6 cells were treated with a series concentration of indicated compounds **11a** and **11b** and infected with SARS-CoV-2 at a multiplicity of infection (MOI) of 0.05. At 24 hours after infection, viral yield in the cell supernatant was quantified by plaque assay. The cytotoxicity of these compounds in Vero E6 cells was also determined by using CCK8 assays. The left and right *y* axis of each graph represents mean percent inhibition of virus yield and mean percent cell viability of the drugs, respectively. (**C** and **D**) Viral RNA copy numbers in the cell supernatants were quantified by qRT-PCR. Data are means ± SD; *n* = 3 biological replicates.

of the interaction are shown in Fig. 3, B and C. The electron density shows that the C of the aldehyde group of **11a** and the catalytic-site Cys^{145} of SARS-CoV-2 M^{pro} form a standard 1.8 Å C–S covalent bond. The oxygen atom of the aldehyde group also plays a crucial role in stabilizing the conformations of the inhibitor by forming a 2.9 Å hydrogen bond with the backbone of residue Cys^{145} in the S1′ site. The (*S*)-γ-lactam ring of **11a** at P1 fits well into the S1 site. The oxygen of the (*S*)-γ-lactam group forms a 2.7 Å hydrogen bond with the side chain of His^{163}. The main chain of Phe^{140} and side chain of Glu^{166} also participate in stabilizing the (*S*)-γ-lactam ring by forming 3.2 Å and 3.0 Å hydrogen bonds with its NH group, respectively. In addition, the amide bonds on the chain of **11a** form hydrogen bonds with the main chains of His^{164} (3.2 Å) and Glu^{166} (2.8 Å), respectively. The cyclohexyl moiety of **11a** at P2 deeply inserts into the S2 site, stacking with the imidazole ring of His^{41}. The cyclohexyl group is also surrounded by the side chains of Met^{49}, Tyr^{54}, Met^{165}, Asp^{187}, and Arg^{188}, producing extensive hydrophobic interactions. The indole group of **11a** at P3 is exposed to solvent (S4 site) and is stabilized by Glu^{166} through a 2.6 Å hydrogen bond. The side chains of residues Pro^{168} and Gln^{189} interact with the indole group of **11a** through hydrophobic interactions. Multiple water molecules (named W1 to W6) play an important role in binding **11a**. W1 interacts with the amide bonds of **11a** through a 2.9 Å hydrogen bond, whereas W2 to W6 contribute to stabilizing **11a** in the binding pocket by forming a number of hydrogen bonds with the aldehyde group of **11a** and the residues of Asn^{142}, Gly^{143}, Thr^{26}, Thr^{25}, His^{41}, and Cys^{44}.

The crystal structure of SARS-CoV-2 M^{pro} in complex with **11b** is very similar to that of the **11a** complex and shows a similar inhibitor-binding mode (Fig. 3D and figs. S3, C and D, and S4A). The difference in binding mode is most probably due to the 3-fluorophenyl group of **11b** at P2. Relative to the cyclohexyl group in **11a**, the 3-fluorophenyl group undergoes a downward rotation (Fig. 3D). The side chains of residues His^{41}, Met^{49}, Met^{165}, Val^{186}, Asp^{187}, and Arg^{188} interact with this aryl group through hydrophobic interactions, and the side chain of Gln^{189} stabilizes the 3-fluorophenyl group with an additional 3.0 Å hydrogen bond (Fig. 3, E and F). In short, these two crystal structures reveal a similar inhibitory mechanism in which both compounds occupy the substrate-binding pocket and block the enzyme activity of SARS-CoV-2 M^{pro}.

Compounds **N1**, **N3**, and **N9** are wide-spectrum inhibitors targeting coronavirus M^{pro}. Compared with the binding modes of **N1**, **N3**, and **N9** in SARS-CoV M^{pro} complex structures reported previously, the binding modes of **11a** and **11b** in SARS-CoV-2 M^{pro} complex structures are similar and the differences among these overall structures are small (Fig. 4 and fig. S4, B to F) (*22*). The differences mainly lie in the interactions at S1′, S2, and S4 subsites, possibly because the sizes of functional groups vary at the corresponding P1′, P2, and P4 sites in the inhibitors (Fig. 4, A and C).

Antiviral activity of 11a and 11b

To further substantiate the enzyme inhibition results, we evaluated the ability of these compounds to inhibit SARS-CoV-2 in vitro (Fig. 5 and fig. S5). As shown in Fig. 5, compounds **11a** and **11b** exhibited good anti–SARS-CoV-2 infection activity in cell culture, with half-maximal effective concentration (EC_{50}) values of 0.53 ± 0.01 μM and 0.72 ± 0.09 μM, respectively, by plaque assay. Neither compound caused cytotoxicity, with half cytotoxic concentration (CC_{50}) values of >100 μM, yielding selectivity indices for **11a** and **11b** of >189 and >139, respectively. We also used immunofluorescence and quantitative real-time polymerase chain reaction (qRT-PCR) to monitor the antiviral activity of **11a** and **11b**. The results showed that **11a** and **11b** exhibit a good antiviral effect on SARS-CoV-2 (Fig. 5 and fig. S5).

Pharmacokinetic and toxicity studies

To explore the further druggability of the compounds **11a** and **11b**, we evaluated both compounds for their pharmacokinetic properties. As shown in table S2, compound **11a** given to mice intraperitoneally (5 mg/kg) and intravenously (5 mg/kg) displayed a half-life ($T_{1/2}$) of 4.27 hours and 4.41 hours, respectively, and we observed a high maximal concentration (C_{max} = 2394 ng/ml) and a good bioavailability of 87.8% when the compound **11a** was given intraperitoneally. Metabolic stability of **11a** in mice was also good (clearance = 17.4 ml min^{-1} kg^{-1}). When administered intraperitoneally (20 mg/kg), subcutaneously (5 mg/kg), and intravenously (5 mg/kg), compound **11b** also showed good pharmacokinetic properties (its bioavailability exceeded 80% when given both intraperitoneally and subcutaneously, and it displayed a longer $T_{1/2}$ of 5.21 hours when given intraperitoneally). Considering the danger of COVID-19, we selected intravenous drip administration for further study because of its comparatively high area under the curve (AUC) value and rapid effect. Relative to **11a** administrated intravenously in CD-1 mice, **11b** displayed a shorter $T_{1/2}$ (1.65 hours) and a faster clearance rate (clearance = 20.6 ml min^{-1} kg^{-1}). Compound **11a** was selected for further investigation with intravenous drip dosing in Sprague-Dawley (SD) rats and beagle dogs. The results showed (table S3) that **11a** exhibited long $T_{1/2}$ values (SD rat, 7.6 hours; beagle dog, 5.5 hours), low clearance rates (rat, 4.01 ml min^{-1} kg^{-1}; dog,

5.8 ml min^{-1} kg^{-1}), and high AUC values (rat, 41,500 hours·ng/ml; dog, 14,900 hours·ng/ml). The above pharmacokinetic results indicate that compound **11a** warrants further study.

An in vivo toxicity study (table S4) of **11a** was carried out on SD rats and beagle dogs. The acute toxicity of **11a** was measured in SD rats. No SD rats died after receiving 40 mg/kg by intravenous drip administration. When the dosage was raised to 60 mg/kg, one of four SD rats died. The dose range toxicity of **11a** was studied for 7 days at dosing levels of 2, 6, and 18 mg/kg in SD rats and at 10 to 40 mg/kg in beagle dogs. All animals received once-daily (QD) dosing by intravenous drip, and all animals were clinically observed at least once a day. No obvious toxicity was observed in either group. These data indicate that **11a** is a good candidate for further clinical study.

REFERENCES AND NOTES

1. N. Zhu *et al*., *N. Engl. J. Med.* **382**, 727–733 (2020).
2. Q. Li *et al*., *N. Engl. J. Med.* **382**, 1199–1207 (2020).
3. J. F. W. Chan *et al*., *Lancet* **395**, 514–523 (2020).
4. P. Zhou *et al*., *Nature* **579**, 270–273 (2020).
5. F. Wu *et al*., *Nature* **579**, 265–269 (2020).
6. R. Lu *et al*., *Lancet* **395**, 565–574 (2020).
7. A. E. Gorbalenya *et al*., bioRxiv 2020.02.07.937862 [preprint]. 11 February 2020.
8. World Health Organization, "WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020" (2020); www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020.
9. B. Cao *et al*., *N. Engl. J. Med.* NEJMoa2001282 (2020).
10. M. L. Holshue *et al*., *N. Engl. J. Med.* **382**, 929–936 (2020).
11. M. Wang *et al*., *Cell Res.* **30**, 269–271 (2020).
12. J. Cohen, *Science* 10.1126/science.abb0659 (27 January 2020).
13. Y. Chen, Q. Liu, D. Guo, *J. Med. Virol.* **92**, 418–423 (2020).
14. S. Hussain *et al*., *J. Virol.* **79**, 5288–5295 (2005).
15. R. Ramajayam, K.-P. Tan, P.-H. Liang, *Biochem. Soc. Trans.* **39**, 1371–1375 (2011).
16. Z. Ren *et al*., *Protein Cell* **4**, 248–250 (2013).
17. K. Anand *et al*., *EMBO J.* **21**, 3213–3224 (2002).
18. H. Yang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13190–13195 (2003).
19. K. Anand, J. Ziebuhr, P. Wadhwani, J. R. Mesters, R. Hilgenfeld, *Science* **300**, 1763–1767 (2003).
20. F. G. Hayden *et al*., *Antimicrob. Agents Chemother.* **47**, 3907–3916 (2003).
21. Y. Kim *et al*., *PLOS Pathog.* **12**, e1005531 (2016).
22. H. Yang *et al*., *PLOS Biol.* **3**, e324 (2005).
23. L. Zhang *et al*., *J. Med. Chem.* acs.jmedchem.9b01828 (2020).
24. Y. Zhai *et al*., *J. Med. Chem.* **58**, 9414–9420 (2015).
25. X. Xue *et al*., *J. Mol. Biol.* **366**, 965–975 (2007).

ACKNOWLEDGMENTS

We thank J. Halpert and LetPub (www.letpub.com) for linguistic assistance during the preparation of this manuscript, and the staff from beamlines BL17U1, BL18U1, and BL19U1 at Shanghai Synchrotron Radiation Facility for assistance during data collection. **Funding:** Supported by National Natural Science Foundation of China grants 21632008, 21672231, 21877118, 31970165, 91953000, and 81620108027; Strategic Priority Research Program of the Chinese Academy of Sciences grants XDA12040107 and XDA12040201; Chinese Academy of Engineering and Ma Yun Foundation grant 2020-CMKYGG-05; Science and Technology Commission of Shanghai Municipality grants 20431900100 and 20431900200; National Key R&D Program of China grants 2017YFC0840300, 2020YFC0841400, 2020YFA0707500, and 2017YFB0202604; Department of Science and Technology of Guangxi Zhuang Autonomous Region grant 2020AB40007; and Frontier Biotechnologies Inc. **Author contributions:** H.Y. and H.L. conceived the project; Y.X., L.-K.Z., H.Y., and H.L. designed the experiments; W.D. and J.L. designed and synthesized the compounds; X.-M.J. and H.S. tested the inhibitory activities; X.X., J.P., C.L., S.H., and J.W. performed the chemical experiments and collected the data; B.Z., Y.Z., Z.J., F.L., F.B., H.W., X.C., X.L., and X.Y. collected the diffraction data and solved the crystal structure; Y.L. and X.C. performed the toxicity experiments; G.X., H.J., Z.R., L.-K.Z., Y.X., H.Y., and H.L. analyzed and discussed the data; and L.-K.Z., Y.X., H.Y., and H.L. wrote the manuscript. **Competing interests:** The Shanghai Institute of Materia Medica has applied for PCT and Chinese patents that cover **11a**, **11b**, and related peptidomimetic aldehyde compounds. **Data and materials availability:** All data are available in the main text or the supplementary materials. The PDB accession number for the coordinates of SARS-CoV-2 M^{pro} in complex with **11a** is 6LZE, and that for the coordinates of SARS-CoV-2 M^{pro} in complex with **11b** is 6M0K. The plasmid encoding the SARS-CoV-2 M^{pro} will be freely available. Compounds **11a** and **11b** are available from H.L. under a material transfer agreement with Shanghai Institute of Materia Medica. There is currently an international effort to join forces to design better inhibitors of SARS-CoV-2 M^{pro} as described in the following website: https://covid.postera.ai/covid. 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/368/6497/1331/suppl/DC1
Materials and Methods
Scheme S1
Figs. S1 to S5
Tables S1 to S4
References (*26*–*29*)

18 March 2020; accepted 20 April 2020
Published online 22 April 2020
10.1126/science.abb4489

INFLUENZA

Different genetic barriers for resistance to HA stem antibodies in influenza H3 and H1 viruses

Nicholas C. Wu[1]*, Andrew J. Thompson[2]*, Juhye M. Lee[3,4,5], Wen Su[6], Britni M. Arlian[2], Jia Xie[7], Richard A. Lerner[7,8], Hui-Ling Yen[6], Jesse D. Bloom[3,4,9], Ian A. Wilson[1,8]†

The discovery and characterization of broadly neutralizing human antibodies (bnAbs) to the highly conserved stem region of influenza hemagglutinin (HA) have contributed to considerations of a universal influenza vaccine. However, the potential for resistance to stem bnAbs also needs to be more thoroughly evaluated. Using deep mutational scanning, with a focus on epitope residues, we found that the genetic barrier to resistance to stem bnAbs is low for the H3 subtype but substantially higher for the H1 subtype owing to structural differences in the HA stem. Several strong resistance mutations in H3 can be observed in naturally circulating strains and do not reduce in vitro viral fitness and in vivo pathogenicity. This study highlights a potential challenge for development of a truly universal influenza vaccine.

The major surface antigen of influenza virus, the hemagglutinin (HA), is composed of a highly variable globular head domain that contains the receptor binding site and a highly conserved stem domain that contains the membrane fusion machinery (*1*). The major antigenic sites on HA are located on the globular head (*2*, *3*), which is immunodominant over the stem (*4*). However, most antibodies to the head domain are strain specific, whereas antibodies to the stem are harder to elicit but have considerably more breadth [reviewed in (*5*)]. The isolation, characterization, and structure determination of broadly neutralizing human antibodies (bnAbs) to the HA stem since 2008 (*6*–*9*) have provided insights into immunogen design toward a universal influenza vaccine (*8*, *10*–*13*) and have offered templates for design of small proteins, peptides, and small molecules against influenza virus (*14*–*18*). Several stem bnAbs are also currently in clinical trials (*19*). Therefore, stem bnAbs open up multiple promising avenues to tackle this challenging global health problem.

Resistance mutations can be a major obstacle for antiviral and vaccine development, but the extent to which this is a problem for stem bnAbs is unclear. Several studies have reported difficulty in selecting strong resistance mutations to stem bnAbs even after extensive virus passaging (*20*–*22*), or through deep mutational scanning (*23*), which is a comprehensive and unbiased approach (*24*). Nonetheless, other studies have obtained strong resistance mutations through virus passaging (*6*, *21*, *25*). Here, we systematically compare the extent to which resistance can emerge to stem bnAbs in HAs from the H3 and H1 subtypes that represent the circulating influenza A strains.

Deep mutational scanning of the major HA stem epitope in H3/HK68 HA

CR9114 (*26*) and FI6v3 (*27*) are two prototypic bnAbs that bind the HA stem and represent those with the greatest breadth. Both FI6v3 and CR9114 neutralize group 1 and 2 influenza A viruses (*26*, *27*), and CR9114 further cross-reacts with influenza B HA (*26*). Deep mutational scanning (*24*), which combines saturation mutagenesis and next-generation sequencing, has been applied to study how HA mutations affect influenza viral fitness (*28*–*30*) and to identify viral mutants that are resistant to antibodies against HA (*31*). Here, we used deep mutational scanning of the HA stem of H3N2 A/Hong Kong/1/1968 (H3/HK68) to search for virus resistance mutations to CR9114 and FI6v3. Unlike previous studies that examined the entire HA (*28*–*31*), we focused on eight residues in HA2 (HA0 is cleaved during maturation into HA1 and HA2) of H3/HK68 HA in and around the main stem epitope (the region recognized by an antibody): namely Q42, I45, D46, Q47, I48, N49, L52, and T111 (Fig. 1, A to C). All except T111 are located on HA2 helix A, which is a common target for stem bnAbs. Residues 42, 45, 46, 48, 49, and 52 were chosen because they interact with CR9114 and FI6v3, and residue 47 because it interacts with FI6v3 (Fig. 1, D and E). Completely buried T111 was selected because its mutation in H5 HA enabled escape from CR6261 (*6*), which binds an epitope similar to CR9114 (*9*, *26*).

By considering mutations to all amino acids at each position, we quantified the in vitro viral fitness of 147 of 152 possible single mutants and 6234 of 10,108 possible double mutants across the eight residues of interest in H3/HK68 under five conditions: no antibody, 2 μg/ml CR9114 immunoglobulin G (IgG), 10 μg/ml CR9114 IgG, 0.3 μg/ml FI6v3 IgG, and 2.5 μg/ml FI6v3 IgG (fig. S1). Many mutants have a relative fitness [proxy for replication fitness (*30*)] similar to that of the wild type (WT), which was set as 1 (Fig. 2, A and B), indicating that the HA stem region can tolerate many mutations.

We further quantified the relative resistance (normalized to WT) of each viral mutant by measuring relative fitness (normalized to WT) with or without antibody (fig. S2). Many resistance mutations to CR9114 and FI6v3 were observed (Fig. 2, C and D), and most were located at HA2 residues 42 and 45, which are important in the binding interface with CR9114 and FI6v3 (Fig. 1, D and E). The double mutants also showed high relative resistance if one mutation exhibited high relative resistance even if the other did not (fig. S3A). Consistently, the relative resistance of double mutants can generally be explained by additivity of their corresponding single mutants (fig. S3B). These results demonstrate the prevalence of H3/HK68 resistance mutations to stem bnAbs and identify cross-resistance mutations to both CR9114 and FI6v3, which are encoded by different germline genes and have very different angles of approach to the HA (*5*).

Validation of resistance mutations

We then individually constructed and tested 18 single and 6 double HA mutants of the H3/HK68 virus spanning a range of relative resistance against CR9114 and FI6v3 IgGs (Fig. 3A). The minimum inhibitory concentration (MIC) in a microneutralization assay strongly correlated with the relative resistance from deep mutational scanning (Spearman's rank correlation > 0.8, Fig. 3B), with several mutants showing strong cross-resistance to both CR9114 and FI6v3. For example, the MICs of CR9114 and FI6v3 against mutants I45Y/S/N/F/W were all >100 and ≥20 μg/ml, respectively, compared to 3.1 and 0.2 μg/ml for WT. This validation experiment substantiates our finding that strong resistance mutations are prevalent in H3/HK68.

Natural occurrence of resistance mutations

Next, we explored whether these resistance mutations can be found in naturally circulating strains and identified a few of these strong resistance mutations at low frequency in human H3N2 isolates (*32*), including I45T, I45M, and N49D (Fig. 4A), although most have not

[1]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [2]Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA. [3]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [4]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. [5]Medical Scientist Training Program, University of Washington, Seattle, WA 98195, USA. [6]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China. [7]Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037, USA. [8]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [9]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: wilson@scripps.edu

been observed yet in naturally circulating strains. I45T is also observed in human H3N2 isolates sequenced without any passaging (fig. S4A), implying that its presence is not due to a passaging artifact (*33*). Moreover, strong cross-resistance mutation I45F was found in all human H2N2 viruses that circulated from 1957 to 1968 (Fig. 4B and fig. S4B), whereas almost all avian H2N2 viruses have Ile[45] (Fig. 4C). This particular mutation in human H2N2 viruses explains why it is more difficult for some stem bnAbs to bind or neutralize this subtype compared to other subtypes, as found in previous studies (*6*, *26*, *34*). Thus, resistance mutations to stem bnAbs already do indeed occur in circulating strains.

In vivo pathogenicity and escape of resistance mutations

Given their relevance to circulating strains, we further tested the in vivo viral pathogenicity of HA2 mutants I45T, I45M, and I45F. The weight loss profiles in mice after infection by HA2 mutant and WT viruses were comparable (fig. S5), indicating that these resistance mutations do not reduce in vivo pathogenicity. We further demonstrated that I45T, I45M, and I45F could escape in vivo prophylactic protection. Whereas mice infected with WT were completely protected by CR9114 IgG at all tested doses (1, 4, and 10 mg/kg), mutants I45T, I45M, and I45F were lethal even at the highest dose of CR9114 IgG (Fig. 4, D to G).

Resistance mutations decrease affinity to bnAbs

To dissect the resistance mechanism, we tested binding of H3/HK68 I45T, I45M, and I45F recombinant HAs to CR9114 and FI6v3, and to another stem bnAb, 27F3 (*34*), which uses the same V_H1-69 germ line as CR9114 and similarly neutralizes group 1 and 2 influenza A viruses. Binding (dissociation constant K_d) of CR9114 Fab, CR9114 IgG, 27F3 Fab, 27F3 IgG, and FI6v3 IgG to the HA mutants (Table 1 and fig. S6) was diminished, particularly with I45F, where binding to CR9114 and 27F3 (Fab and IgG) was undetectable. By contrast, binding of these stem Fabs and IgGs to the HA mutant N49T, which did not exhibit resistance against CR9114 and FI6v3 (Figs. 2C and 3A), was comparable to that of the WT (Table 1). As a control, we also tested binding of bnAb S139/1 that targets the receptor binding site far from the stem epitope (*35*, *36*). S139/1 IgG affinities for those HA mutants (K_d = 1.8 to 3.1 nM) were similar to that of the WT (K_d = 2.1 nM). Thus, virus resistance to stem bnAbs correlated with a decrease in binding affinity for the mutant HAs.

To evaluate the structural basis of the resistance, we determined crystal structures of

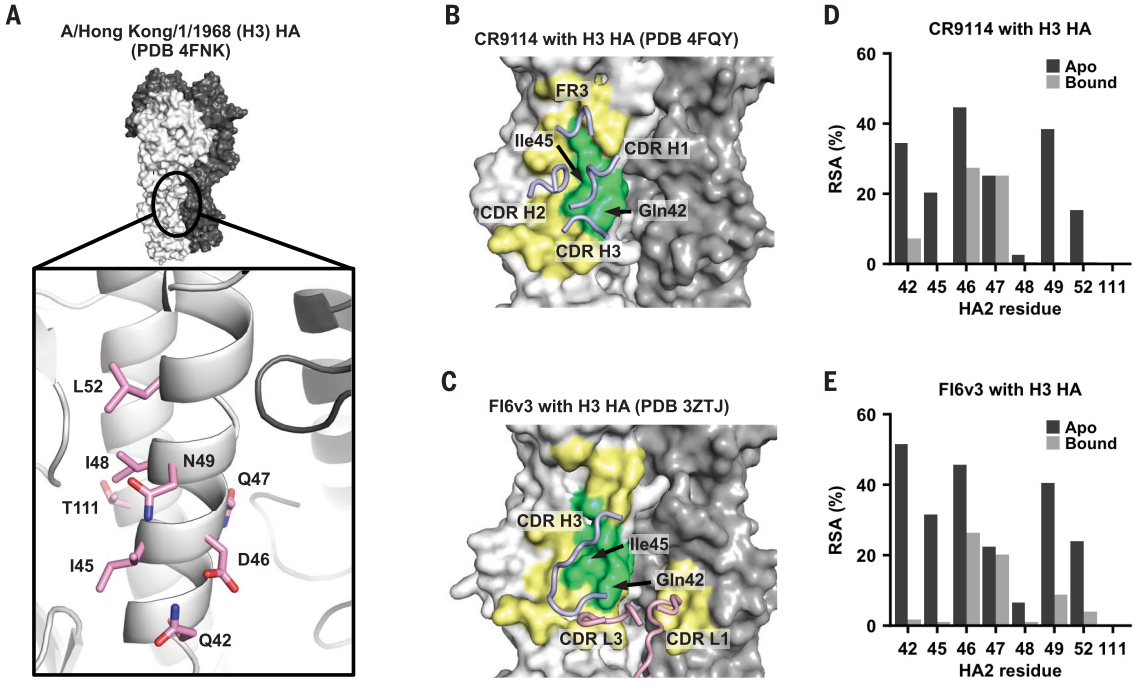


Fig. 1. Epitopes of broadly neutralizing antibodies to the HA stem. (**A**) The location of residues of interest in this study on the HA structures. All residues of interest are on HA2. One protomer of the trimer is shown in light gray and the other two protomers in dark gray, and a detailed view of the location of the residues of interest is shown in the inset. (**B** and **C**) Epitopes of (B) CR9114 Fab in complex with H3 HA (PDB 4FQY) (*26*) and (C) FI6v3 in complex with H3 HA (PDB 3ZTJ) (*27*) are colored in yellow and green, with residues of interest colored in green. The arrows indicate the positions of HA2 residues 42 and 45, which are in the center of the bnAb epitopes. Antibody paratopes (CDRs and FR regions) are shown in tube representation and labeled accordingly. Blue, heavy chain; pink, light chain. (**D** and **E**) The relative solvent accessibility (RSA) of each residue of interest is shown. Black bar, apo form; gray bar, Fab-bound form.

Table 1. Binding affinity of IgGs and Fabs against HAs of WT and different H3/HK68 HA2 mutants. n.b., no binding.

K_d (nM)	WT	HA2-I45T	HA2-I45M	HA2-I45F	HA2-N49T
CR9114 Fab	43.2 ± 0.4	779.6 ± 50.0	319.1 ± 12.1	n.b.	11.3 ± 0.4
27F3 Fab	122.6 ± 3.0	n.b.	n.b.	n.b.	105.0 ± 1.0
CR9114 IgG	<0.1	164.8 ± 2.3	75.0 ± 1.0	n.b.	<0.1
27F3 IgG	1.7 ± 0.1	n.b.	n.b.	n.b.	1.9 ± 0.9
FI6v3 IgG	<0.1	7.8 ± 0.2	6.7 ± 0.1	62.0 ± 0.2	<0.1
S139/1 IgG	2.1 ± 0.1	3.1 ± 0.3	1.9 ± 0.1	1.8 ± 0.1	2.8 ± 0.1

HAs with HA2 mutations I45T, I45M, and I45F to 2.1 to 2.5 Å resolutions (table S1 and fig. S7A). Compared to WT (Ile[45]), the shorter side chain of I45T would create a void when CR9114 is bound (fig. S7B) that would be energetically unfavorable. By contrast, the longer flexible side chain of I45M would likely clash with CR9114 (fig. S7B), but CR9114 can still bind the I45M mutant, albeit with much lower affinity than WT (Table 1). The I45F mutant, however, would clash more severely with CR9114, and no binding was detected (Table 1 and fig. S7B). Similar observations for FI6v3 (fig. S7C) explain the sensitivity of CR9114 and FI6v3 to mutations at HA2 residue 45.

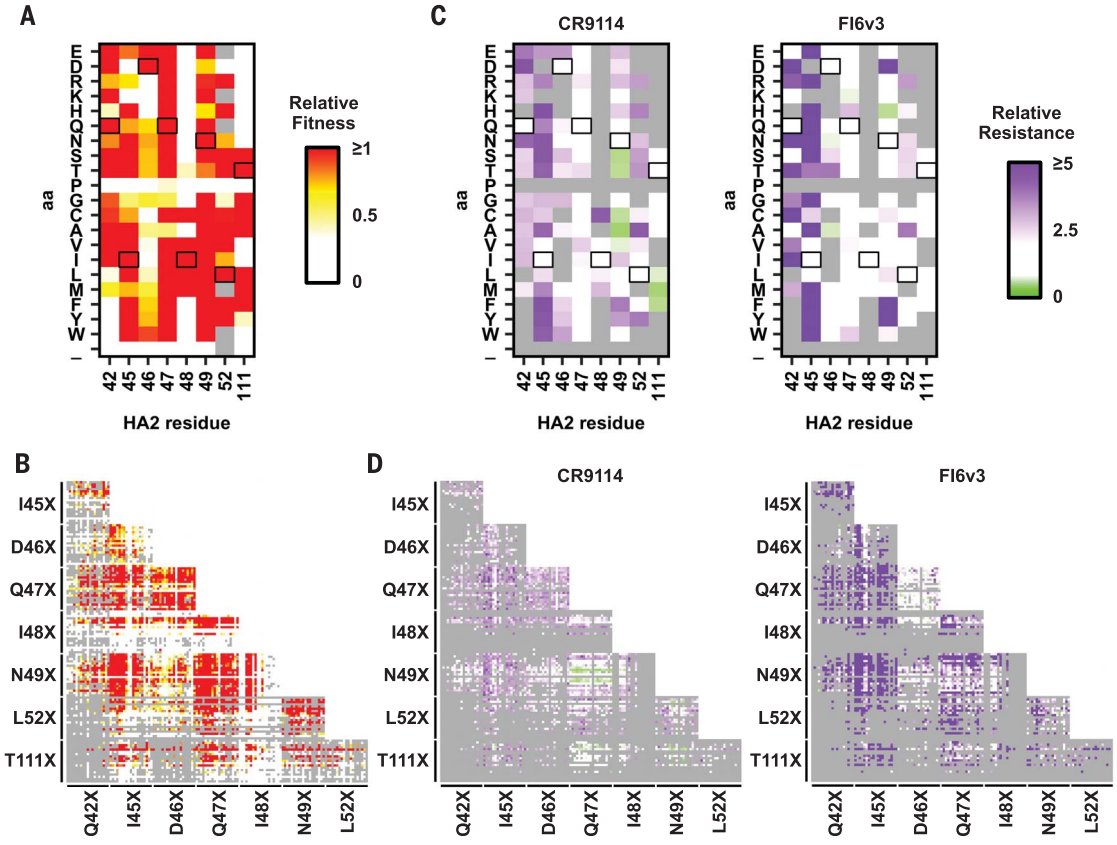


Fig. 2. Fitness and resistance profile of H3/HK68 HA2 single and double viral mutants. (**A** and **B**) On the basis of the deep mutational scanning experiment, the relative fitness of (A) single and (B) double mutants is shown with WT set as 1. In (A) and (B), mutants with a next-generation sequencing read count of less than 20 from the plasmid mutant library are excluded and shown as gray. (**C** and **D**) Relative resistance for (C) each single and (D) each double mutant against 10 μg/ml CR9114 antibody or 2.5 μg/ml FI6v3 antibody is shown. Relative resistance for WT is set as 1. In (C) and (D), mutants with a relative fitness of less than 0.5 are shown as gray. Residues that correspond to the WT sequence are boxed. In (B) and (D), each color point in the heatmap represents a double mutant. Of note, some mutants with an increased sensitivity to antibody are shown in green on the relative resistance color scale. 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.

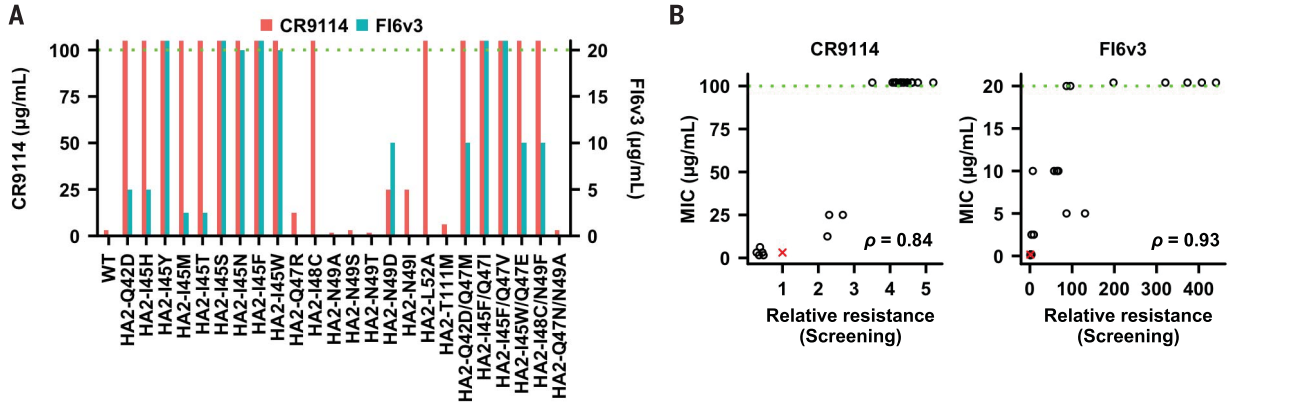


Fig. 3. Characterization of antibody-resistant mutants. (**A**) The minimum inhibitory concentration (MIC) of CR9114 and FI6v3 to individual viral mutants is shown. The MIC of CR9114 is in red and represented by the *y* axis on the left. The MIC of FI6v3 is in blue and represented by the *y* axis on the right. (**B**) The Spearman's rank correlations (ρ) between the MIC measured from individual mutants and the relative resistance (against 10 μg/ml CR9114 antibody or 2.5 μg/ml FI6v3 antibody) computed from the deep mutational scanning experiment (screening) are shown. The green dashed line represents the upper detection limit in (A) and (B). Wild type is represented by the red "X".

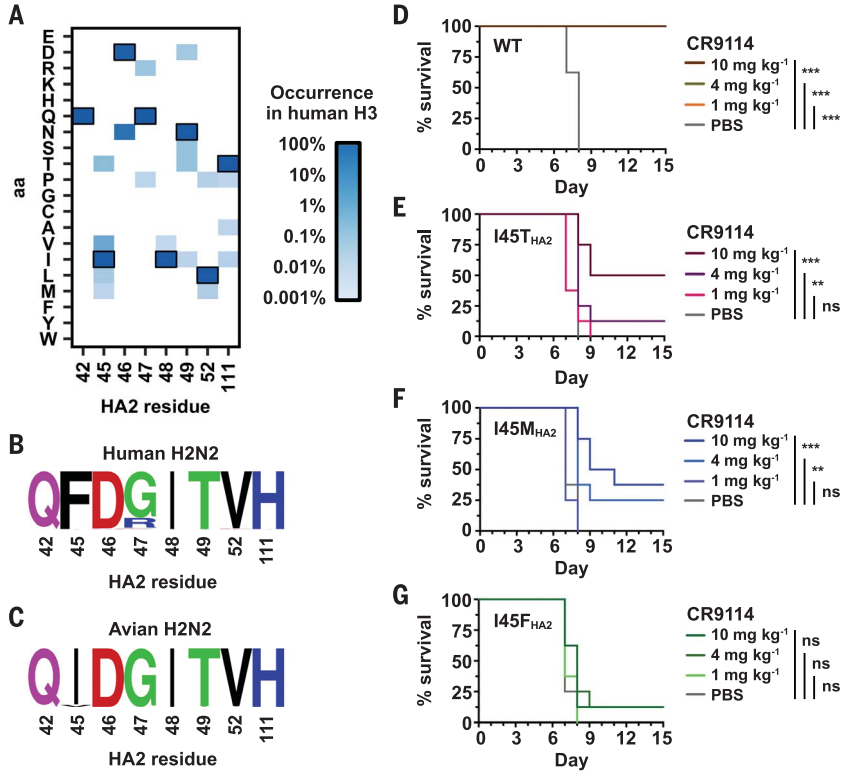


Fig. 4. In vivo characterization and natural occurrence of antibody-resistant mutants. (**A**) The natural occurrence frequencies of different amino acid variants at the residues of interest in human H3 HAs are shown as a heatmap. (**B** and **C**) The natural occurrence frequencies of different amino acid variants at the residues of interest in (B) human H2N2 HA or (C) avian H2N2 HA are shown as sequence logos. (**D** to **G**) Prophylactic protection experiments were performed with different doses of CR9114 against (D) WT, (E) HA2 I45T mutant, (F) HA2 I45M mutant, and (G) HA2 I45F mutant. Recombinant H3/HK68 (7:1 on H1/PR8 backbone) viruses were used. Lethal doses [25 median lethal dose (MLD_{50})] of WT or mutant viruses were used. Kaplan-Meier survival curves are shown. Paired analysis of each treatment group, relative to control, was conducted using Log-rank (Mantel-Cox) tests. ***$P \leq 0.001$; **$P \leq 0.01$; ns (not significant), $P > 0.05$.

Resistance to HA stem bnAbs is subtype specific

We further examined whether mutations that conferred strong resistance in H3/HK68 would have the same phenotypes in other H3 strains. Consequently, we examined I45T, I45M, and I45F mutations on H3N2 A/Wuhan/359/95 (H3/Wuhan95) virus. These mutants had WT-like titers after viral rescue and passaging (fig. S8A), showed a plaque size similar to that of WT (fig. S8B), conferred strong resistance to FI6v3 (fig. S8, C and D), and, therefore, have similar phenotypes in H3/HK68 and H3/Wuhan95 viruses. This result led us to hypothesize that strong resistance mutants to stem bnAbs are readily attainable in human H3 strains that span a wide range of years, as well as explore whether the same phenomenon can be observed in the H1 subtype, which is the other currently circulating influenza A subtype in the human population.

In contrast to our findings for H3/HK68, deep mutational scanning found that resistance mutants of H1N1 A/WSN/33 (H1/WSN) virus to FI6v3 were rare and had only very small effects (*23*). To further investigate the difference between H3 and H1 strains, we performed four additional deep mutational scanning experiments—three with H1N1 strains, namely A/Solomon Islands/3/2006 (H1/SI06) against FI6v3, A/Michigan/45/2015 (H1/Mich15) against FI6v3, and H1/WSN against CR9114; and one with H3N2 strain A/Perth/16/2009 (H3/Perth09) against FI6v3. The H1/SI06 and H1/Mich15 HA mutant virus libraries contain all possible single substitutions at HA2 residues 42, 45, 46, 47, 48, 49, 52, and 111, whereas H1/WSN and H3/Perth09 HA mutant virus libraries both contain all possible single substitutions across the entire HA and were constructed previously (*37*, *38*). We also analyzed the previously published dataset on H1/WSN against FI6v3 (*23*).

To compare H1/SI06, H1/Mich15, H1/WSN, and H3/Perth09 to H3/HK68, we computed the relative resistance of mutations at HA2 residues 42, 45, 46, 47, 48, 49, 52, and 111 (Fig. 5, A to E). Similar to H3/HK68, resistance mutations are highly prevalent in H3/Perth09 (Fig. 5C). Conversely, resistance mutations were rare in H1/SI06, H1/Mich15, and H1/WSN (Fig. 5, A to E). We further calculated the fraction surviving (*23*) for each viral mutant across the entire H1/WSN and H3/Perth09 HAs during antibody selection. Fraction surviving is a quantitative measure for the resistance and is normalized across deep mutational scanning experiments (*23*). For example, fraction surviving values of 1, 0.5, and 0.1 indicate that the replication fitness in the presence of antibody selection is 100, 50, and 10% of that without antibody selection. The fraction surviving values of H1/WSN mutants against CR9114 were all very small, similar to previous observations of H1/WSN against FI6v3 (Fig. 5F and fig. S10). In stark contrast, many mutants of H3/Perth09 were identified with a large fraction surviving value (Fig. 5F). Consistent with the relative resistance profile of H3/HK68 (Fig. 2C), a number of H3/Perth09 mutants with a large fraction surviving value were found at HA2 residues 42 and 45 (Fig. 5F and fig. S11A). Moreover, mutations at HA2 residue 53, which were not examined in H3/HK68 (Fig. 2), had high fraction surviving values in H3/Perth09 against FI6v3 (fig. S11, A and B). In H3 HA, mutation of HA2 residue 53 would abolish a hydrogen bond to FI6v3 (fig. S11B). Together, these results suggest that the prevalence of resistance mutations to stem bnAbs is a general phenomenon for the H3 subtype but not the H1 subtype.

Subtype-specific differences in the HA stem

We next aimed to elucidate the mechanism that underlies the lower genetic barrier to resistance to stem bnAbs in H3 HA as compared to H1 HA. Many mutations at HA2 residue 45 have a high fitness cost in H1/SI06 (fig. S9A), which can increase the genetic barrier to resistance. However, most mutations at HA2 residue 45 have no fitness cost in H1/Mich15. In addition, the mutational fitness profiles of H1/SI06 and H1/Mich15 (fig. S9, A and B) show that many mutations can be tolerated in the HA stem, similar to H3/HK68 (Fig. 2A and fig. S9, C to F). Thus, the difference in genetic barrier to resistance to stem bnAbs between H1 and H3 subtypes cannot be fully explained by their ability to tolerate mutations (i.e., fitness cost of mutations).

We therefore further compared the structures of FI6v3 in complexes with H1 HA and

H3 HA (*27*). The orientation of Tyr100c [Kabat numbering (*39*)] on the third complementarity determining region of the antibody heavy chain (CDR H3) of FI6v3 differs when binding to H1 or H3 HAs (Fig. 5G). The position of HA2 Ile^{45} also differs between H1 and H3 HAs when FI6v3 is bound. As a result, Tyr100c of FI6v3 packs tighter to HA2 Ile^{45} of H3 than of H1 HA. Thus, a bulkier substitution at residue 45 of HA2 will disrupt binding of FI6v3 to H3 HA more than to H1 HA. Therefore, the low genetic barrier to resistance to stem bnAbs in the H3 subtype can be partly attributed to both high mutational tolerance in the HA stem and subtype-specific structural features, but it is likely that the latter plays a more critical role in determining subtype-specific differences. Although a number of subtype-specific structural features are known in the stem region (*40*), how they influence the genetic barriers for resistance to stem bnAbs remains to be addressed in future studies.

Implications for escape from a universal vaccine or therapeutic stem bnAbs

Prior studies of influenza bnAbs have not considered whether different subtypes might have different abilities to generate resistance mutations against bnAbs. We find here that H3 HA has a much lower genetic barrier to resistance to two of the broadest bnAbs, CR9114 and FI6v3, as compared to H1 HA, consistent with reports of strong resistance to other human HA stem antibodies in the H3 subtype (*21*, *25*, *41*) versus none (*21*) to weak resistance (*22*, *23*) in the H1 subtype. Therefore, it may be easier for stem bnAbs to suppress H1 rather than H3 subtype viruses.

Because the HA stem is immunosubdominant to the head domain, immunological pressure on the HA stem may not yet be sufficient to affect the evolution of circulating human viruses (*42*). However, several stem bnAbs are currently in clinical trials (*19*), and some recently developed influenza vaccine immunogens have focused on targeting the HA stem (*5*, *8*). If stem bnAbs begin to be distributed on a global scale, the immunological pressure on the HA stem will increase. Our findings here indicate that resistance mutations could emerge, at least in the H3 subtype.

Although resistance mutations to stem bnAbs are still rare in currently circulating influenza strains (Fig. 4A), it is important to evaluate the potential impact of such mutations, because many vaccine strategies aim to elicit anti-stem antibodies. We could not overcome some key resistance mutations (I45T, I45M, and I45F) by in vitro evolution of CR9114 (fig. S12). Nonetheless, the best strategy to prevent or overcome such resistance may involve delivery or elicitation of a combination of antibodies with different resistance profiles. The discovery and characterization of bnAbs with

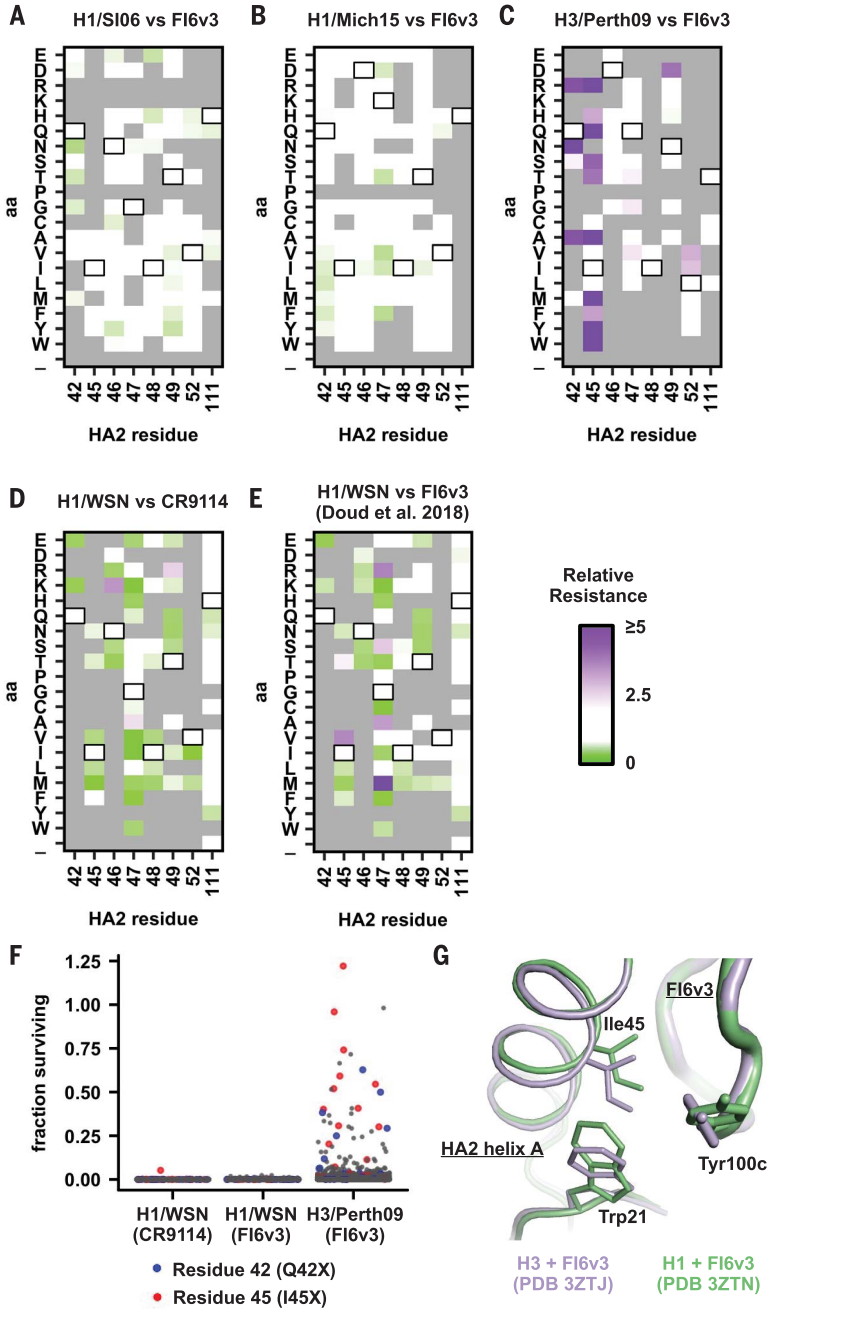


Fig. 5. Relative resistance profile of single mutants in multiple strains of influenza virus. (**A** to **E**) Relative resistance is measured for each single viral mutant at HA2 residues 42, 45, 46, 47, 48, 49, 52, and 111 of (A) H1/SI06 against 300 ng/ml FI6v3 antibody, (**B**) H1/Mich15 against 300 ng/ml FI6v3 antibody, (**C**) H1/WSN against 100 ng/ml CR9114 antibody, (**D**) H1/WSN against 200 ng/ml FI6v3 antibody [data are from (*23*)], and (**E**) H3/Perth09 against 15 μg/ml FI6v3 antibody. Relative resistance for WT is set as 1. Mutants with a relative fitness of less than 0.5 are excluded and shown as gray. Residues that correspond to the WT sequence are boxed. (**F**) Fraction surviving for all single mutants across the HA protein is shown. Each data point represents one mutant. Fraction surviving was computed as previously described (*23*). Assuming no antibody-mediated enhancement of virus replication, the theoretical upper limit for the fraction surviving is 1, which indicates that the replication fitness is the same with and without antibody selection (in practice, fraction surviving values slightly >1 can sometimes be obtained because of the experimental error in quantitative polymerase chain reaction and next-generation sequencing). Data for H1/WSN against FI6v3 were from a previous study (*23*). Data points that represent mutations at residues 42 and 45 are colored in blue and red, respectively, and mutations at other residues are in gray. (**G**) The crystal structures of H3 HA in complex with FI6v3 (PDB 3ZTJ) (*27*) and H1 HA in complex with FI6v3 (PDB 3ZTN) (*27*) were compared by aligning their heavy-chain variable domains. Because the structure of CR9114 with H1 HA is not available, a similar analysis was not performed on CR9114.

different escape profiles will continue to be key to broadening our arsenal against influenza virus. For example, human H2N2 virus, which carries a Phe at HA2 residue 45, often has low reactivity with stem bnAbs (*6*, *26*, *27*, *34*, *43*), although a few can have high potency against human H2N2 (*44*–*46*). Future studies on anti-stem responses against human H2N2 and emerging viruses, such as H5N1 and H7N9, may provide further insights into how to overcome potential resistance should immune pressure become focused on the HA stem.

REFERENCES AND NOTES

1. I. A. Wilson, J. J. Skehel, D. C. Wiley, *Nature* **289**, 366–373 (1981).
2. D. C. Wiley, I. A. Wilson, J. J. Skehel, *Nature* **289**, 373–378 (1981).
3. A. J. Caton, G. G. Brownlee, J. W. Yewdell, W. Gerhard, *Cell* **31**, 417–427 (1982).
4. F. Krammer, P. Palese, *Nat. Rev. Drug Discov.* **14**, 167–182 (2015).
5. N. C. Wu, I. A. Wilson, *J. Mol. Biol.* **429**, 2694–2709 (2017).
6. M. Throsby *et al*., *PLOS ONE* **3**, e3942 (2008).
7. J. Sui *et al*., *Nat. Struct. Mol. Biol.* **16**, 265–273 (2009).
8. N. C. Wu, I. A. Wilson, *Nat. Struct. Mol. Biol.* **25**, 115–121 (2018).
9. D. C. Ekiert *et al*., *Science* **324**, 246–251 (2009).
10. A. Impagliazzo *et al*., *Science* **349**, 1301–1306 (2015).
11. H. M. Yassine *et al*., *Nat. Med.* **21**, 1065–1070 (2015).
12. S. A. Valkenburg *et al*., *Sci. Rep.* **6**, 22666 (2016).
13. K. S. Corbett *et al*., *mBio* **10**, e02810–e02818 (2019).
14. S. J. Fleishman *et al*., *Science* **332**, 816–821 (2011).
15. T. A. Whitehead *et al*., *Nat. Biotechnol.* **30**, 543–548 (2012).
16. A. Chevalier *et al*., *Nature* **550**, 74–79 (2017).
17. R. U. Kadam *et al*., *Science* **358**, 496–502 (2017).
18. M. J. P. van Dongen *et al*., *Science* **363**, eaar6221 (2019).
19. E. Sparrow, M. Friede, M. Sheikh, S. Torvaldsen, A. T. Newall, *Vaccine* **34**, 5442–5448 (2016).
20. G. Nakamura *et al*., *Cell Host Microbe* **14**, 93–103 (2013).
21. N. Chai *et al*., *PLOS Pathog.* **12**, e1005702 (2016).
22. C. S. Anderson *et al*., *Sci. Rep.* **7**, 14614 (2017).
23. M. B. Doud, J. M. Lee, J. D. Bloom, *Nat. Commun.* **9**, 1386 (2018).
24. D. M. Fowler, S. Fields, *Nat. Methods* **11**, 801–807 (2014).
25. R. H. Friesen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 445–450 (2014).
26. C. Dreyfus *et al*., *Science* **337**, 1343–1348 (2012).
27. D. Corti *et al*., *Science* **333**, 850–856 (2011).
28. N. C. Wu *et al*., *Sci. Rep.* **4**, 4942 (2014).
29. B. Thyagarajan, J. D. Bloom, *eLife* **3**, e03300 (2014).
30. N. C. Wu *et al*., *Cell Host Microbe* **21**, 742–753.e8 (2017).
31. M. B. Doud, S. E. Hensley, J. D. Bloom, *PLOS Pathog.* **13**, e1006271 (2017).
32. R. B. Squires *et al*., *Influenza Other Respir. Viruses* **6**, 404–416 (2012).
33. N. C. Wu *et al*., *PLOS Pathog.* **13**, e1006682 (2017).
34. S. Lang *et al*., *Cell Rep.* **20**, 2935–2943 (2017).
35. R. Yoshida *et al*., *PLOS Pathog.* **5**, e1000350 (2009).
36. P. S. Lee *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17040–17045 (2012).
37. M. B. Doud, J. D. Bloom, *Viruses* **8**, 155 (2016).
38. J. M. Lee *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8276–E8285 (2018).
39. Kabat numbering scheme is commonly used standard for numbering the amino-acid residues in an antibody. Third complementarity-determining region of heavy chain (CDR H3) ranges from residues 95 to 102 in the Kabat numbering scheme. Residues inserted in CDR H3 are numbered with 100a, 100b, 100c, and so on.
40. S. J. Zost, N. C. Wu, S. E. Hensley, I. A. Wilson, *J. Infect. Dis.* **219** (suppl. 1), S38–S45 (2019).
41. D. C. Ekiert *et al*., *Science* **333**, 843–850 (2011).
42. E. Kirkpatrick, X. Qiu, P. C. Wilson, J. Bahl, F. Krammer, *Sci. Rep.* **8**, 10432 (2018).
43. M. G. Joyce *et al*., *Cell* **166**, 609–623 (2016).
44. Y. Okuno, Y. Isegawa, F. Sasao, S. Ueda, *J. Virol.* **67**, 2552–2558 (1993).
45. C. Dreyfus, D. C. Ekiert, I. A. Wilson, *J. Virol.* **87**, 7149–7154 (2013).
46. N. L. Kallewaard *et al*., *Cell* **166**, 596–608 (2016).
47. J. Lee, J. Bloom, jbloomlab/HA_stalkbnAb_MAP: journal submission, Version journal_submission, Zenodo (9 February 2020); https://doi.org/10.5281/zenodo.3660467.
48. N. C. Wu, wchnicholas/HAstemEscape: publication, Version publication, Zenodo (9 February 2020); https://doi.org/10.5281/zenodo.3660739.
49. N. C. Wu, wchnicholas/CR9114mut: publication, Version publication, Zenodo (9 February 2020); https://doi.org/10.5281/zenodo.3660731.

ACKNOWLEDGMENTS

We thank W. Yu and G. Grande for technical support in protein expression; S. Head, J. Ledesma, and P. Natarajan (TSRI Next Generation Sequencing Core and Fred Hutch Genomics Core) for next-generation sequencing; M. Haynes and B. Seegers (TSRI Flow Cytometry Core Facility) for performing FACS; A. Ho, E.-M. Strauch, and B. Graham for insightful discussions; and J. Paulson for his continual support. **Funding:** We acknowledge support from the Bill and Melinda Gates Foundation OPP1170236 (to I.A.W.), NIH K99 AI139445 (to N.C.W.), F30 AI136326 (to J.M.L.), R01 AI127893 (to J.D.B.), R56 AI127371 (to I.A.W.), and R01 AI114730 (to J. C. Paulson) This work was partially supported through an NIAID Collaborative Influenza Vaccine Innovation Centers (CIVIC) contract (75N93019C00051). J.M.L. was supported in part by the Center for Inference and Dynamics of Infectious Diseases (CIDID), funded by NIH U54 GM111274. J.D.B. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** N.C.W., J.M.L., R.A.L., H.-L.Y., J.D.B., and I.A.W. conceived and designed the study. N.C.W. and J.M.L. performed the deep mutational scanning experiments. N.C.W., J.M.L., and J.D.B. performed the computational data analysis. N.C.W. performed the structural analysis and yeast display experiment. N.C.W. and W.S. performed functional characterization of the viral mutants. A.J.T. and B.M.A. performed the in vivo characterization of the viral mutants. N.C.W. and J.X. produced the CR9114 and 27F3 antibodies. J.M.L. produced the FI6v3 antibody. N.C.W. and I.A.W. wrote the paper, and all authors reviewed and edited the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw sequencing data have been submitted to the NIH Short Read Archive under accession numbers BioProject PRJNA510654, PRJNA493101, and PRJNA510700. X-ray coordinates and structure factors are deposited at the RCSB Protein Data Bank under accession codes 6NHP, 6NHQ, and 6NHR. Computer codes and processed data are deposited at Zenodo (*47*–*49*). All other data that support the conclusions are included in the manuscript or supplementary materials. The pHW2000 plasmids encoding the eight genes of A/Wuhan/359/95 (H3N2) virus and HA plasmids containing the HA2 mutations are available for noncommercial use from H.-L.Y. under a material transfer agreement (University of Hong Kong). All other materials are available from I.A.W. on request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1335/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S3
References (*50*–*69*)

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

16 September 2019; accepted 14 April 2020
10.1126/science.aaz5143

BIODIVERSITY CHANGE

Landscape-scale forest loss as a catalyst of population and biodiversity change

Gergana N. Daskalova[1]*, Isla H. Myers-Smith[1], Anne D. Bjorkman[2,3], Shane A. Blowes[4,5], Sarah R. Supp[6], Anne E. Magurran[7], Maria Dornelas[7]

Global biodiversity assessments have highlighted land-use change as a key driver of biodiversity change. However, there is little empirical evidence of how habitat transformations such as forest loss and gain are reshaping biodiversity over time. We quantified how change in forest cover has influenced temporal shifts in populations and ecological assemblages from 6090 globally distributed time series across six taxonomic groups. We found that local-scale increases and decreases in abundance, species richness, and temporal species replacement (turnover) were intensified by as much as 48% after forest loss. Temporal lags in population- and assemblage-level shifts after forest loss extended up to 50 years and increased with species' generation time. Our findings that forest loss catalyzes population and biodiversity change emphasize the complex biotic consequences of land-use change.

Accelerating human impacts are reshaping Earth's ecosystems (*1*). The abundance of species' populations (*2*, *3*) and the richness (*4*–*6*) and composition (*6*) of ecological assemblages at sites around the world are being altered over time in complex ways (*7*–*9*). However, there is currently only a limited quantitative understanding of how global change drivers, such as land-use change, influence the observed heterogeneous local-scale patterns in population abundance and biodiversity (*8*, *10*, *11*). In terrestrial ecosystems, much current knowledge stems from space-for-time approaches (*12*, *13*) and model projections (*14*, *15*) that attribute population and richness declines to different types of land-use change, including reductions in forest cover. Yet space-for-time methods may not accurately represent the effects of global change drivers, because they do not account for ecological lags (*8*, *16*, *17*) and community self-regulation (*18*). Furthermore, ongoing controversy about the diverse impacts of habitat fragmentation on biodiversity (*19*–*21*) could be in part attributable to a lack of observational data from sites encompassing the full spectrum of forest fragmentation. Recent global-scale datasets of past land cover reconstructions (*22*) and contemporary high-resolution remote-sensing observations (*23*, *24*) provide an opportunity to quantify landscape-scale decreases and increases in forested areas around the world (hereafter, forest loss and gain). By integrating forest loss estimates with population and biodiversity observations (*25*, *26*) (Figs. 1 and 2A), our analysis provides insight into the influence of land-use change on local-scale population and biodiversity change around the planet.

In our study, we set out to conduct a global-extent attribution analysis of the influence of forest cover change on population and biodiversity change (Fig. 1). We quantitatively tested specific predictions of how the timing and magnitude of landscape-scale forest loss influence species' populations and ecological assemblages across terrestrial ecosystems around the planet (Figs. 1 and 2 and table S1) (*27*). Land-use change, and particularly forest cover loss, alters habitat and resource availability (*12*, *28*, *29*) and is a global threat for the persistence of terrestrial species (*30*) (Fig. 2 and fig. S12). We thus predicted the greatest impacts on populations and biodiversity when time-series monitoring encompassed the 10-year period that included the largest reduction in forested areas at each site (calculated between the years 850 and 2015, hereafter "all-time peak forest loss"). We also expected greater population and species richness declines and higher turnover after, relative to before, contemporary peak forest loss (i.e., the year of the largest reduction in forested area within the duration of each time series). Finally, species with longer generation times typically respond more slowly to environmental change (*31*). We thus predicted that lags in ecological responses to forest loss would increase with longer generation times across taxa.

We measured landscape-scale historic and contemporary forest loss by integrating information from the Land Use Harmonization (*22*) and Global Forest Change (*23*) databases. We also used the ESA Landcover (*32*) and KK09 (*33*) databases to examine whether our results were consistent across land-use change data sources. We compared historic and contemporary forest loss with temporal population change (trends in the numerical abundance of species) and biodiversity change (trends in species richness and turnover in assemblage composition) (Figs. 1 and 2). We analyzed 2729 populations of 730 species and biodiversity change in 3361 ecological assemblages (Figs. 2A and 3). We measured population change using the Living Planet Database, which includes 133,092 records of the number of individuals of a species in a given area over time (*25*). We measured biodiversity change using the BioTIME database, which comprises 4,970,128 records of the number and abundance of species in ecological assemblages over time (*26*). Together, these time series represent a range of taxa including amphibians (388), birds (5090), mammals (266), reptiles (76), invertebrates (80), and plants (187) and 2157 sites that cover almost the entire spectrum of forest loss and gain around the world (Fig. 2B). We used a standardized cell size of approximately 96 km^2 to match response variables (population change, richness change, and turnover) to landscape-scale forest change. Our analyses were robust to the spatial scale over which we calculated forest change (figs. S13 and S14) (*27*).

We carried out the following workflow for our global assessment of the consequences of forest cover change for population and biodiversity trends over time. To relate population and biodiversity change to historic forest loss, we quantified the baseline all-time peak forest loss at each site. To relate population and biodiversity change to contemporary forest loss, we compared population and biodiversity change before and after contemporary peak forest loss. To investigate temporal lags, we quantified the time period between contemporary peak forest loss and maximum change in populations and assemblages detected after peak forest loss had occurred at each site (Fig. 1B). We calculated population change (μ) using state-space models that account for observation error and random fluctuations (*34*), and calculated richness change (slopes of rate of change over time) using mixed-effects models. We quantified temporal change in species composition as the turnover component of Jaccard's dissimilarity measure (change due to species replacement) (*35*). Turnover is often independent of changes in species richness (*9*) and is the dominant component of compositional change across time series of ecological assemblages (*36*). We used a hierarchical Bayesian modeling framework, with individual time series nested within biomes (*37*), to account for the spatial structure of the data (*27*).

Historical baselines

In line with our first prediction ("historical baselines"), we found that local-scale population declines were more pronounced when the monitoring occurred during the period

[1]School of GeoSciences, University of Edinburgh, Edinburgh EH9 3FF, Scotland. [2]Biological and Environmental Sciences, University of Gothenburg, 405 30 Gothenburg, Sweden. [3]Gothenburg Global Biodiversity Centre, 405 30 Gothenburg, Sweden. [4]German Centre for Integrative Biodiversity Research (iDiv), 04103 Leipzig, Germany. [5]Department of Computer Science, Martin Luther University Halle-Wittenberg, 06108 Halle (Salle), Germany. [6]Data Analytics Program, Denison University, Granville, OH 43023, USA [7]Centre for Biological Diversity, University of St Andrews, St Andrews KY16 9TF, Scotland.
*Corresponding author. Email: gndaskalova@gmail.com

of all-time peak forest loss (Fig. 1B and Fig. 3, B and C). For many of the sites represented by the time series we studied, major changes in forest cover occurred during the past two centuries. In regions such as Europe and North America, all-time peak forest loss was typically during the early 1800s, before population, biodiversity, and satellite monitoring had begun (Figs. 2C and 3B). The Living Planet Database and BioTIME time series captured more than half of the spectrum of contemporary forest cover change around the world, in contrast to previous criticisms of some of these data underrepresenting areas with anthropogenic impact (*38*) (Fig. 2, B and C, and Fig. 3B). Yet in only ~5% of monitored time series did forest loss lead to a conversion in the dominant habitat type (e.g., from primary forest to urban areas). Habitat conversions corresponded with both gains and losses in populations and biodiversity, with the highest rates of turnover when primary forests were converted to agricultural and urban areas or to secondary forests (fig. S17). The links among historical baselines, the timing of all-time peak forest loss, and resulting ecological change emphasize the need for a long-term perspective to quantify the complexity of biodiversity change in the Anthropocene (*11*, *17*).

Contemporary forest loss

Contrary to our second prediction ("contemporary forest loss"), we found that forest loss acted as a catalyst amplifying both increases and decreases in local-scale populations and assemblages over time (Figs. 3 and 4 and figs. S4 to S6, S9, and S10). Across time series, more than half of all populations and assemblages (61%) experienced higher rates of change after the largest forest loss event within each time series. Contemporary peak forest loss intensified population declines, population increases, and richness losses—but not richness gains—relative to the period before peak forest loss (Fig. 4). In nearly one-third of time series (32%), more than 10% of the species in the assemblage at the time of contemporary peak forest loss were replaced by new species by the end of the time series (Fig. 4, G and H). The assemblages that experienced the most richness change also experienced the most turnover [Pearson correlation = 0.37; 95% confidence interval (CI) = 0.31 to 0.43]. The influence of contemporary peak forest loss on population and biodiversity change was not strongly correlated to the magnitude of the specific forest loss event (figs. S4 to S6). Our findings indicate a wide spectrum of population and biodiversity responses to forest loss that might be overlooked without accounting for temporal dynamics and lagged responses (*12*, *13*, *15*, *39*).

Temporal lags

In line with our third prediction ("temporal lags"), we found evidence for up to half-century ecological lags in local-scale changes in population abundance, species richness, and turnover after contemporary peak forest loss (Fig. 5). On average, we documented maximum change in populations and ecological assemblages 6 to 13 years after forest loss across taxa. Yet nearly half of population and biodiversity change (40%) happened within 3 years of peak forest loss, demonstrating that rapid shifts in populations and assemblages occur frequently after habitat change (Fig. 5 and fig. S7). Consistent with our prediction, the

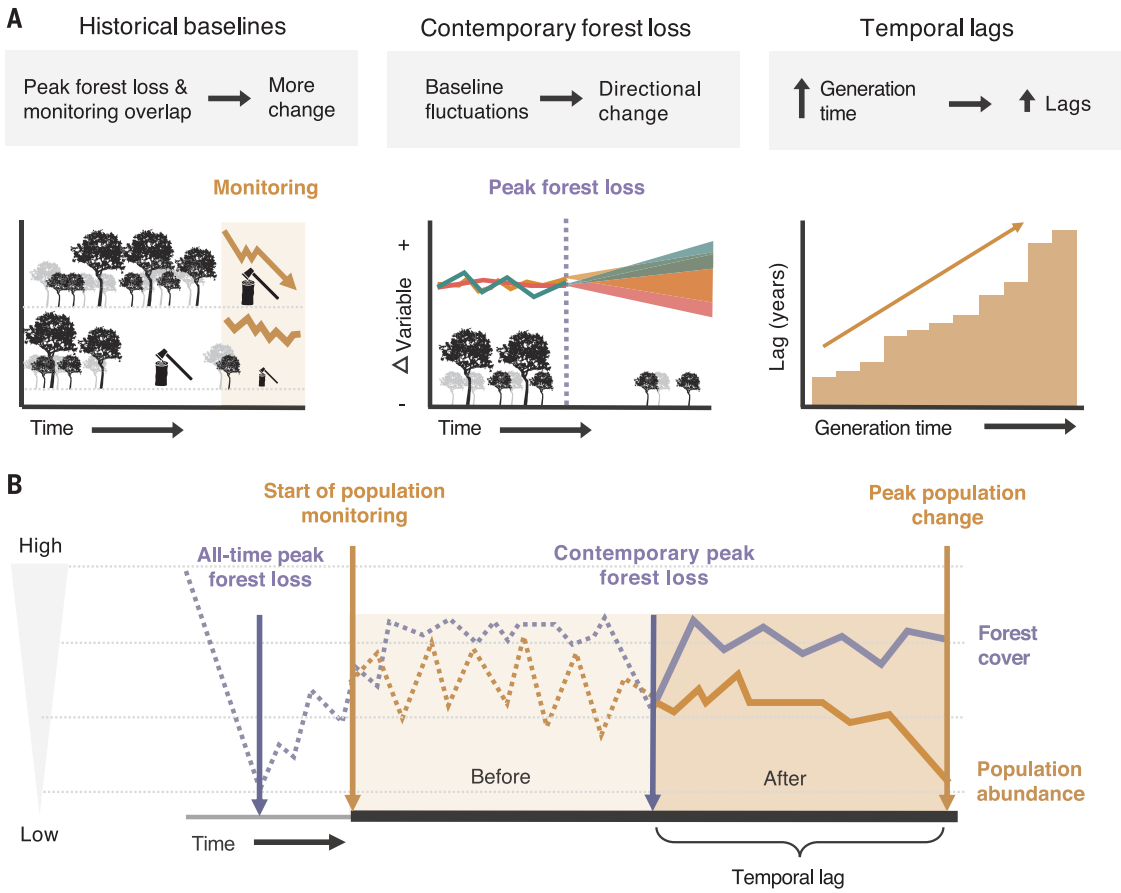


Fig. 1. Influence of forest loss on population and biodiversity change. We tested three pathways through which forest loss can influence the population abundance of species and the richness and turnover (temporal species replacement) of ecological assemblages: historical baselines of forest loss, timing of contemporary forest loss, and temporal lags in population and biodiversity responses. (**A**) Conceptual diagram of our predictions outlined with respect to population change, richness change, and turnover. (**B**) Analytical workflow for determining all-time and contemporary peak forest loss and temporal lags [further details in (*27*)].

period between peak forest loss and peak change in populations and biodiversity was longer for taxa with longer generation times (e.g., large mammals and birds; Fig. 5B and table S2). Population declines and increases occurred on similar time scales (Fig. 5C). Losses in species richness lagged behind gains by approximately half a year (slope = 0.5, CI = 0.1 to 1.05), indicating that extinction debts and immigration credits accumulated at roughly the same speed across taxa. The similar pace and temporal delay of population declines and increases, and of richness gains and losses, could help to explain previous findings of community self-regulation (*18*) and no net population

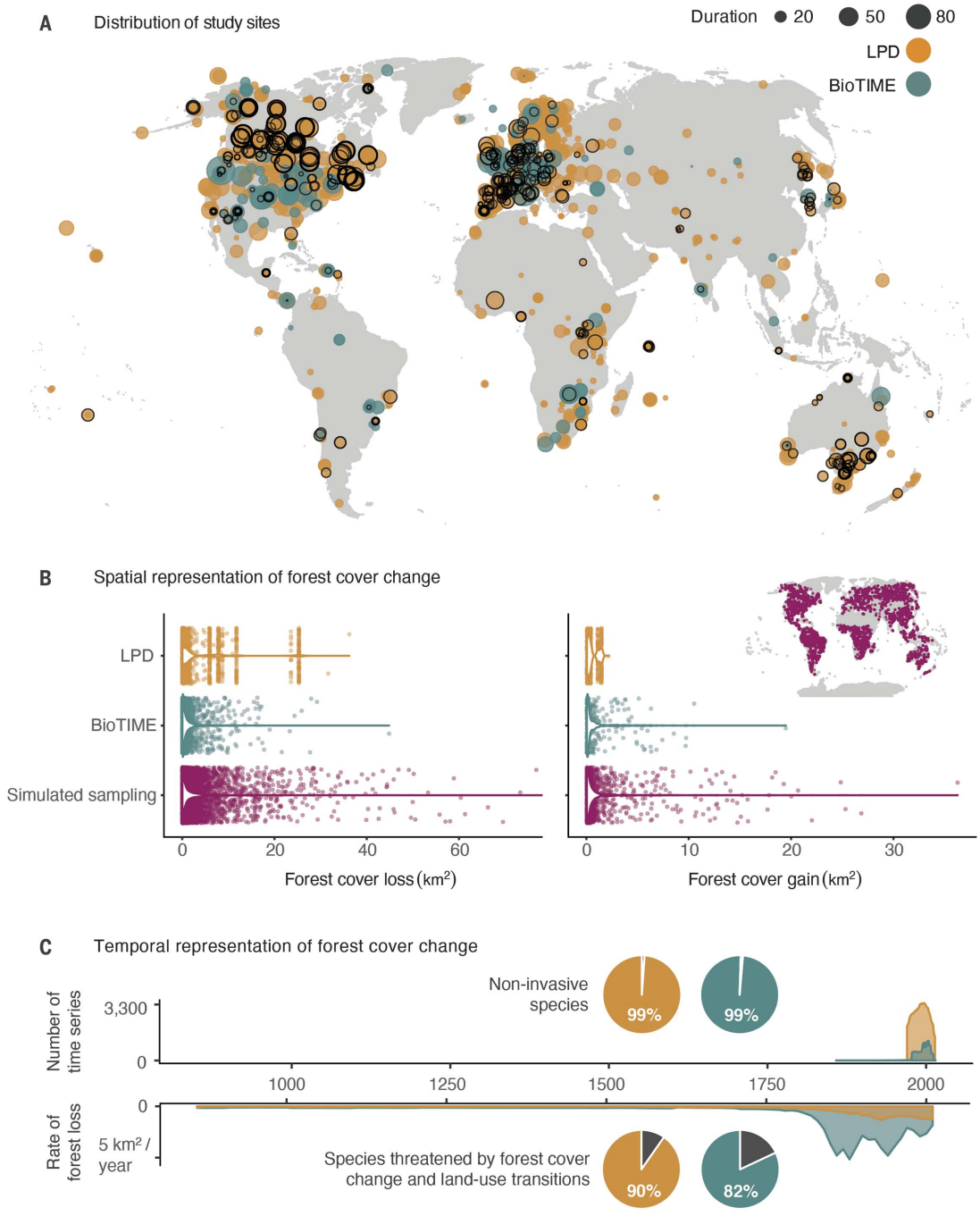


Fig. 2. Population and biodiversity monitoring over time broadly spans the global variation in forest cover change. (**A**) Locations and duration of 542 Living Planet Database (LPD) and 199 BioTIME studies, containing 6090 time series from 2157 sites [black outline denotes sites that were forested at the start of the monitoring (1247 sites); see table S1 for sample size in each woody biome]. (**B**) Among all time series, 44% experienced historic or contemporary forest loss of comparable magnitude to forest cover change across a simulated random sample of geographical locations (shown on map inset) from the global distribution of forest cover loss and gain. We did not detect directional effects of the magnitude of forest gain across monitored sites (figs. S4 to S6). (**C**) The number of time series increased over time (top), but the rates of forest loss were often higher before the start of monitoring (bottom; see figs. S2 and S3 for variation in monitoring periods among time series). Insets show the proportion of study species that are not classified as invasive (top) and that are threatened by land-use change, based on species' IUCN threat assessments (bottom; see fig. S12 for details).

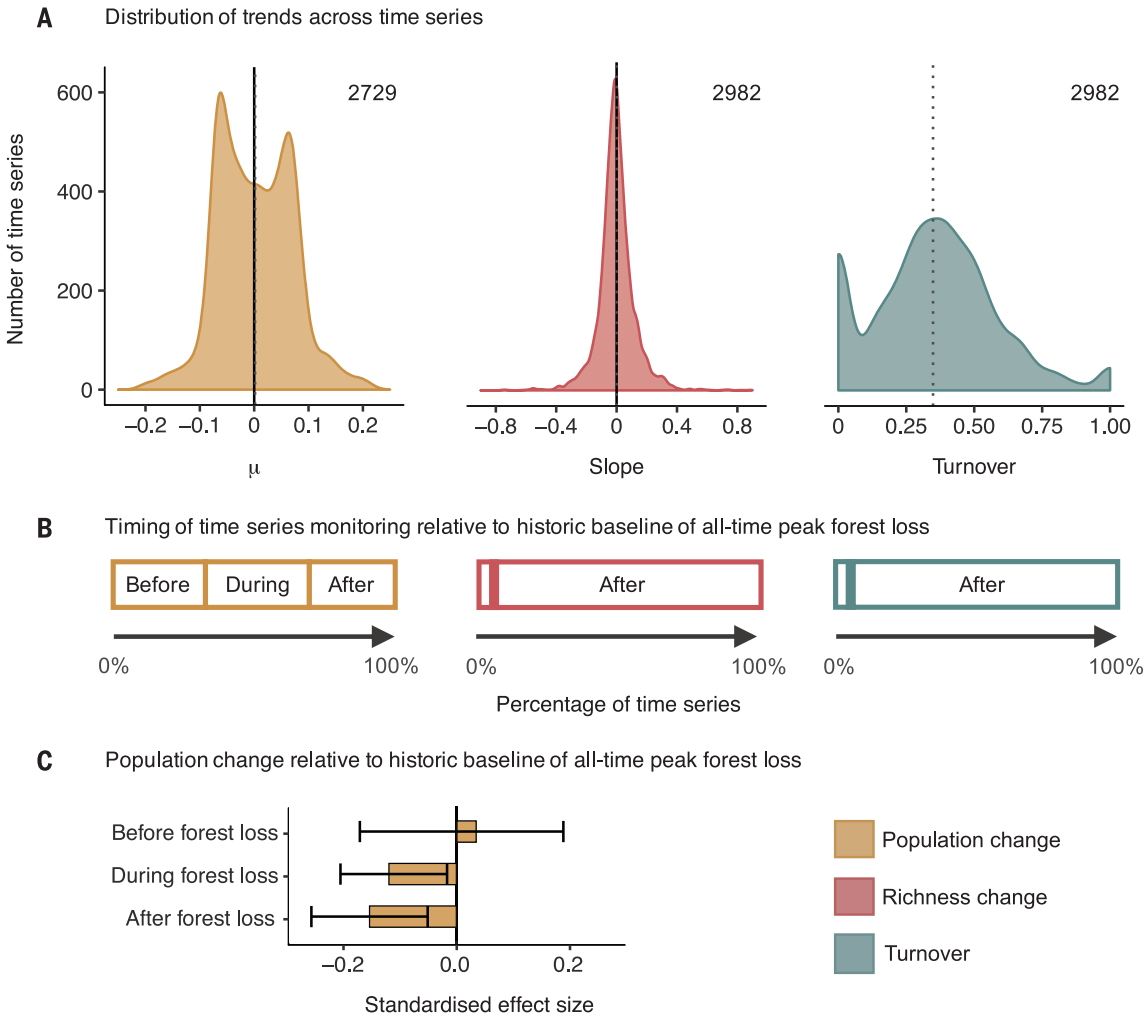


Fig. 3. Heterogeneity in population and biodiversity trends and land-use histories from sites around the world. (**A**) All three metrics of ecological change (population change, richness change, and turnover) show heterogeneous distributions across sites. (**B**) Population monitoring occurred at different time periods relative to all-time peak forest loss (for 33% of sites before, for 37% during, and for 30% after), whereas biodiversity monitoring predominantly started after all-time peak forest loss had occurred (94% of sites). (**C**) Population declines were more acute when all-time peak forest loss occurred during the population monitoring period (slope = –0.007, CI = –0.012 to –0.001; see table S2 for model outputs). Low sample size for the "before" (101) and "during" (38) categories precluded a similar analysis for richness change and turnover. Lines in (A) denote zero (solid) and mean values (dotted). Numbers in (A) show sample size (i.e., number of time series).

change (*2*, *3*, *10*) or richness change (*5*, *6*) at local scales. Temporal lags in biodiversity change have also been observed in post-agricultural forests (*4*, *40*) and fragmented grasslands (*31*), where agricultural activity ceased decades to centuries ago, yet richness and assemblage composition change continue to the modern day. Overall, our results indicate that increasing rates of land-use change in the Anthropocene (*41*, *42*) will alter ecosystems on both short- and long-term time scales that need to be captured in ongoing and future biodiversity monitoring.

Heterogeneity in responses to forest cover loss could be due to a number of factors, including (i) temporal lags in population or assemblage responses, as observed in our study and elsewhere (*17*, *31*); (ii) context-specific responses to forest loss, such as the same amount of habitat change corresponding to biodiversity declines at one site but increases at another (*13*, *43*, *44*); and (iii) interactions with other drivers occurring simultaneously with forest loss (*11*, *45*, *46*). Our finding that forest loss was concurrent with both declines and increases in populations and assemblages is consistent with the varied and often positive effects of habitat fragmentation on biodiversity metrics such as species richness (*19*). However, forest loss occurring outside of the period of population or biodiversity monitoring, as well as the type of woody vegetation being gained and lost, might influence our ability to detect a causal link between forest loss and biodiversity change (*17*, *47*). Increases in woody vegetation caused by agroforestry or plantations might not reflect ecosystem recovery, such as with natural succession after forest cover loss (*48*–*50*). Our finding that forest cover gain did not directly correspond with gains in population abundance and species richness highlights the need for high-resolution temporal data of the specific vegetation types constituting forest cover changes around the world. The complexity and heterogeneity of forest cover change effects on biodiversity (*13*, *43*, *51*, *52*) demonstrate that caution is warranted with respect to recent calls for global afforestation as a climate change mitigation tool (*53*).

Variation in species' vulnerability to forest cover loss (*43*, *51*) may be contributing to the wide spectrum of population and biodiversity responses to shifts in forest cover. Species that have experienced frequent habitat disturbance during their evolutionary history might be more resilient to land-use change, whereas novel habitat alterations could have a greater influence on species' persistence and abundance (*13*, *43*) (Fig. 3). In a post hoc test, we found that in forest-dominated sites, where past disturbances were likely less frequent, declines in species' abundance were more frequent than increases when contemporary forest loss occurred, whereas richness change and turnover did not show directional trends (fig. S16). Additionally, in our study, rare and common species, as defined by their range size, mean population size, and habitat specificity (*54*), responded in similar ways to forest loss (figs. S11 and S12). In contrast to this result, space-for-time comparisons that do not account for temporal dynamics and lagged responses have found that land-use change has a more negative impact on rare species

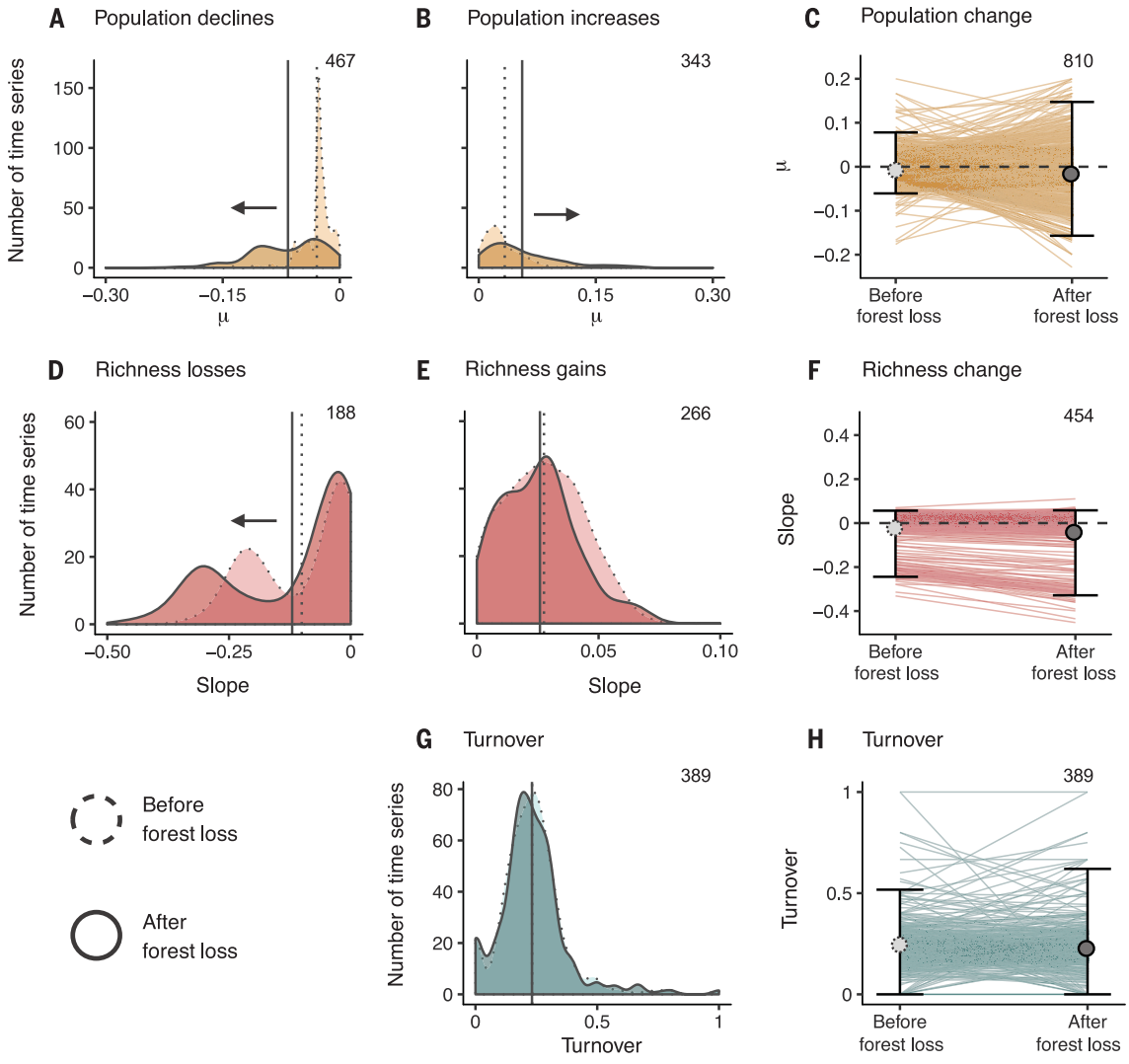


Fig. 4. At the site level, population and biodiversity change increase after contemporary peak forest loss. In total, population and richness change increased across 61% and decreased across 39% of the 1653 time series for which baseline comparisons were possible (i.e., the time series were long enough to include at least 5 years before and after forest loss). Only turnover included instances of no difference in the amount of change before and after peak forest loss (6% of time series). (**A**, **B**, **D**, **E**, and **G**) Distributions compare population declines (μ) (A), population increases (μ) (B), richness losses (slopes) (D), richness gains (slopes) (E), and turnover (Jaccard's dissimilarity) (G) in the periods before and after contemporary peak forest loss, the largest forest loss event during the monitoring of each site. Vertical lines over distributions show the mean for each category (dotted, before; solid, after). (**C**, **F**, and **H**) Temporal trends before and after peak forest loss are indicated with lines for individual time series. Light and dark gray points and error bars show mean values and 2.5% and 97.5% quantiles. Duration varied among time series but was consistent for each individual time series (i.e., *n* years before forest loss = *n* years after forest loss, *n* ≥ 5 years; see fig. S8 for relationship between duration and number of survey points). Numbers on plots indicate sample size. See table S2 for model outputs.

than on common species (*55*). Accounting for both inter- and intraspecific heterogeneity in species' vulnerability to forest cover change is key when scaling from localized impacts of human activities to global-scale biodiversity patterns and attribution of change (*1*, *19*–*21*, *39*, *43*, *51*).

Taxonomic, spatial, and temporal imbalances in sampling can make large-scale attribution analyses of biodiversity trends and global change drivers challenging and can influence the inferences that we draw from such studies (figs. S2, S3, S8, S9, and S11 to S14). For this reason, we explored in greater detail three specific challenges of our terrestrial biodiversity attribution analyses. First, tropical species and locations are underrepresented in current open-source temporal biodiversity databases (Fig. 2A) (*38*). In a post hoc test, we found that in the tropics, where there is intense, often unprecedented forest loss, the effects of forest loss were stronger and more negative across sites with available data, relative to the rest of the globe (figs. S9 and S10 and tables S1 and S2). Second, the spatial scales at which biodiversity is monitored (from 1 m^2 to 25 × 10^8 km^2) and the resolution of forest cover datasets (from 30 m to ~20 km; figs. S13 and S14) could introduce spatial mismatches between the driver and response. Nonetheless, we found that the heterogeneous relationships among richness change, turnover, and forest loss were consistent across forest loss calculated on scales from 10 to 500 km^2 (fig. S14). Third, temporal mismatches and lags (Figs. 1C and 5) can obscure the ways in which forest loss may be related to population and biodiversity change. We found that attribution signals were strongest when a peak in forest loss occurred during the time-series monitoring (Figs. 3 and 4). Our results indicate that biodiversity assessments and global change attribution analyses will be improved by better spatial and temporal matching of biodiversity and environmental impact data.

Our analysis reveals an intensification of both increases and decreases of populations and biodiversity by up to 48% after forest loss at sites around the planet. This finding demonstrates heterogeneity in the influence of forest cover change on populations and ecological assemblages and challenges the assumption that land-use change predominantly leads to population declines and species richness loss (*12*, *14*, *39*). A current assumption underlying existing projections of biodiversity responses to land-use change (*12*, *14*) is that space-for-time approaches accurately reflect longer-term population and biodiversity dynamics (*41*). In contrast, we found temporal

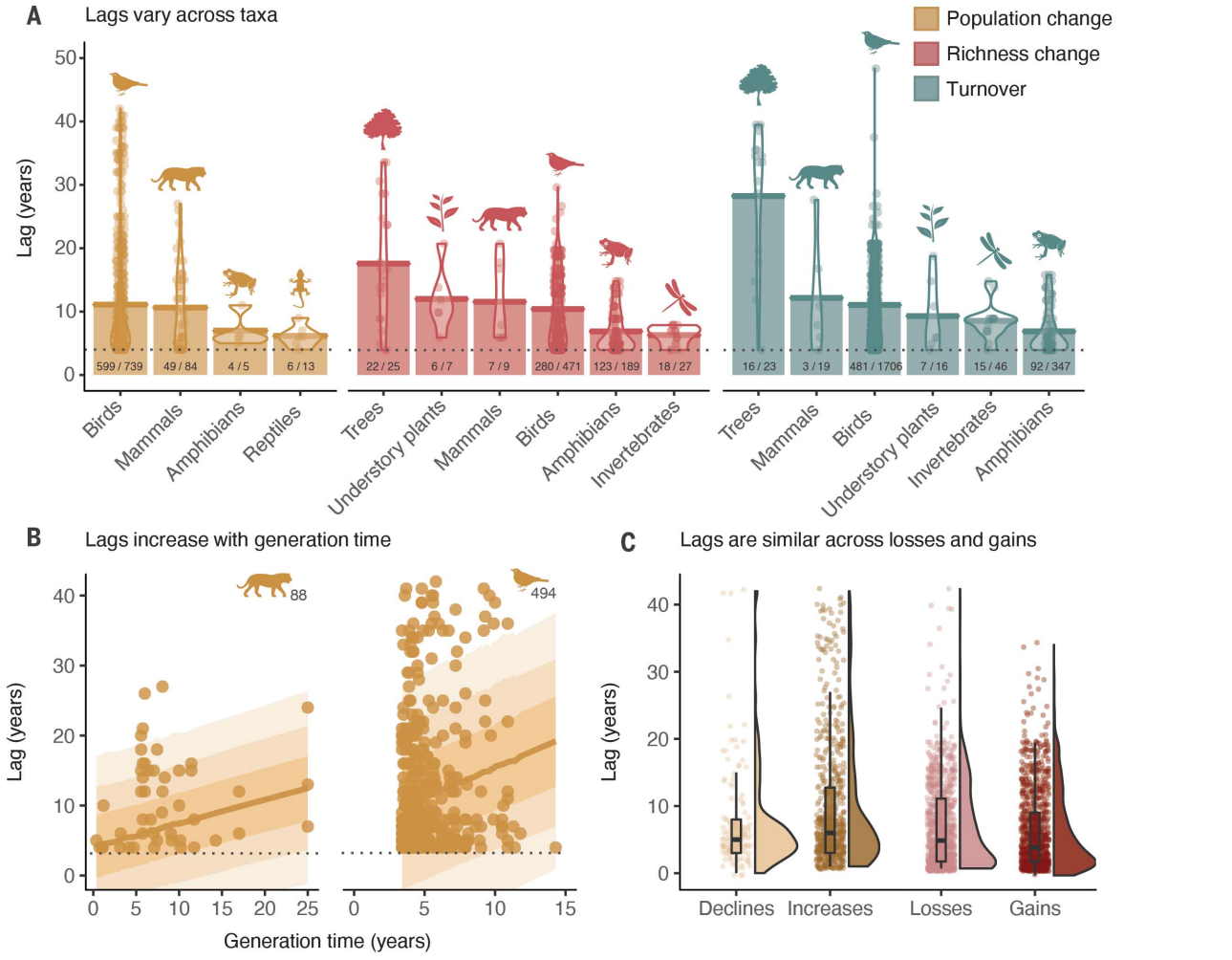


Fig. 5. Temporal lags in population and biodiversity change after contemporary peak forest loss. Population and assemblage change after contemporary peak forest loss may be delayed by up to half a century, with taxa and species with long generation times showing the longest temporal lags. (**A**) We compared lags of 3 (dashed horizontal line) or more years between peak forest loss during the monitoring for each time series and peak change in population or biodiversity across six taxa (sample size was 841 time series for population change, 728 for richness change, and 2157 for turnover). Bars show mean lag for each taxon, violins show the distribution of lag values, and the points are lag values for each time series. Numbers on bars indicate how many time series experienced lags out of the total sample size for each taxon. (**B**) We found that temporal lags in mammal and bird population change increased with longer species generation times. (**C**) Temporal lags were similar across population declines and increases and across species richness losses and gains. See table S2 for model outputs.

lags of up to 50 years in population and biodiversity change after forest loss that differed across taxa and among species' generation times. Our analyses highlight that the local-scale responses of populations and assemblages to forest cover loss and gain are complex and variable over time. Incorporating the full spectrum of population and biodiversity responses to land-use change will improve projections of the future impacts of global change on biodiversity and thus contribute to the conservation of the world's biota during the Anthropocene.

REFERENCES AND NOTES

1. IPBES, Summary for Policymakers of the *Global Assessment Report on Biodiversity and Ecosystem Services* of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (2019).
2. G. N. Daskalova, I. H. Myers-Smith, J. L. Godlee, bioRxiv 272898 [preprint]. 16 April 2018.
3. M. Dornelas *et al.*, *Ecol. Lett.* **22**, 847–854 (2019).
4. L. Baeten, M. Hermy, S. Van Daele, K. Verheyen, *J. Ecol.* **98**, 1447–1453 (2010).
5. M. Vellend *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19456–19459 (2013).
6. M. Dornelas *et al.*, *Science* **344**, 296–299 (2014).
7. A. E. Magurran *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1843–1847 (2018).
8. N. G. Yoccoz, K. E. Ellingsen, T. Tveraa, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1681–1683 (2018).
9. H. Hillebrand *et al.*, *J. Appl. Ecol.* **55**, 169–184 (2018).
10. B. Leung, D. A. Greenberg, D. M. Green, *Divers. Distrib.* **23**, 1372–1380 (2017).
11. D. E. Bowler *et al.*, *People Nat.* **2**, 380–394 (2020).
12. T. Newbold *et al.*, *Nature* **520**, 45–50 (2015).
13. M. G. Betts *et al.*, *Nature* **547**, 441–444 (2017).
14. T. Newbold, *Proc. R. Soc. B* **285**, 20180792 (2018).
15. T. Newbold, D. P. Tittensor, M. B. J. Harfoot, J. P. W. Scharlemann, D. W. Purves, bioRxiv 439059 [preprint]. 10 October 2018.
16. S. C. Elmendorf *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 448–452 (2015).
17. J.-B. Mihoub *et al.*, *Sci. Rep.* **7**, 41591 (2017).
18. N. J. Gotelli *et al.*, *Sci. Adv.* **3**, e1700315 (2017).
19. L. Fahrig, *Annu. Rev. Ecol. Evol. Syst.* **48**, 1–23 (2017).
20. N. M. Haddad *et al.*, *Ecography* **40**, 48–55 (2017).
21. E. I. Damschen *et al.*, *Science* **365**, 1478–1480 (2019).
22. G. C. Hurtt *et al.*, *Clim. Change* **109**, 117–161 (2011).
23. M. C. Hansen *et al.*, *Science* **342**, 850–853 (2013).
24. S. Channan, K. Collins, W. R. Emanuel, *Global Mosaics of the Standard MODIS Land Cover Type Data* (University of Maryland and Pacific Northwest National Laboratory, 2014); https://modis.gsfc.nasa.gov/data/dataprod/mod12.php.
25. Living Planet Index database (2016); www.livingplanetindex.org.
26. M. Dornelas *et al.*, *Glob. Ecol. Biogeogr.* **27**, 760–786 (2018).
27. See supplementary materials.
28. R. Elahi *et al.*, *Curr. Biol.* **25**, 1938–1943 (2015).
29. D. F. Sax, S. D. Gaines, *Trends Ecol. Evol.* **18**, 561–566 (2003).
30. IUCN Red List of Threatened Species, Version 2017-3 (2017); www.iucnredlist.org.
31. J. Krauss *et al.*, *Ecol. Lett.* **13**, 597–605 (2010).
32. ESA Climate Change Initiative, ESA Land Cover Product (1992-2015). ESA Climate Change Initiative—Land Cover led by UCLouvain (2017); www.esa-landcover-cci.org/?q=node/175.
33. J. O. Kaplan, K. M. Krumhardt, N. Zimmermann, *Quat. Sci. Rev.* **28**, 3016–3034 (2009).

34. J.-Y. Humbert, L. S. Mills, J. S. Horne, B. Dennis, *Oikos* **118**, 1940–1946 (2009).
35. A. Baselga, *Glob. Ecol. Biogeogr.* **19**, 134–143 (2010).
36. S. A. Blowes *et al*., *Science* **366**, 339–345 (2019).
37. D. M. Olson, E. Dinerstein, *Ann. Mo. Bot. Gard.* **89**, 199 (2002).
38. A. Gonzalez *et al*., *Ecology* **97**, 1949–1960 (2016).
39. G. Ceballos, P. R. Ehrlich, R. Dirzo, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6089–E6096 (2017).
40. M. Vellend *et al*., *Ecology* **87**, 542–548 (2006).
41. A. De Palma *et al*., *Adv. Ecol. Res.* **58**, 163–199 (2018).
42. L. Egli, C. Meyer, C. Scherber, H. Kreft, T. Tscharntke, *Glob. Change Biol.* **24**, 2212–2228 (2018).
43. M. G. Betts *et al*., *Science* **366**, 1236–1239 (2019).
44. M. G. Betts, B. Phalan, S. J. K. Frey, J. S. Rousseau, Z. Yang, *Divers. Distrib.* **24**, 439–447 (2018).
45. F. E. B. Spooner, R. G. Pearson, R. Freeman, *Glob. Change Biol.* **24**, 4521–4531 (2018).
46. J. D. Fridley, J. P. Wright, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4702–4706 (2018).
47. F. Isbell, D. Tilman, P. B. Reich, A. T. Clark, *Nat. Ecol. Evol.* **3**, 1533–1538 (2019).
48. J. W. Veldman *et al*., *Science* **366**, eaay7976 (2019).
49. P. Potapov *et al*., *Ecol. Soc.* **13**, art51 (2008).
50. P. G. Curtis, C. M. Slay, N. L. Harris, A. Tyukavina, M. C. Hansen, *Science* **361**, 1108–1111 (2018).
51. C. D. L. Orme *et al*., *Nat. Ecol. Evol.* **3**, 886–891 (2019).
52. C. Banks-Leite *et al*., *Science* **345**, 1041–1045 (2014).
53. J.-F. Bastin *et al*., *Science* **365**, 76–79 (2019).
54. D. Rabinowitz, in *The Biological Aspects of Rare Plants Conservation*, H. Synge, Ed. (Wiley, 1981), pp. 205–217.
55. L. Sykes, L. Santini, A. Etard, T. Newbold, *Conserv. Biol.* **34**, 688–696 (2019).

ACKNOWLEDGMENTS

We thank WWF International and the Zoological Society of London for compiling the Living Planet Database; R. Freeman and L. McRae for useful discussions; the BioTIME team for compiling the BioTIME database; F. Moyes for managing the BioTIME database; the creators of the Land Use Harmonization Database; the Hansen lab for producing the Forest Cover Change Database; NASA for producing the MODIS Landcover Database; the Forest & Nature Lab at Ghent University for a stimulating discussion on historic and contemporary land-use change and choosing appropriate baselines for comparison of biodiversity change through time; A. Phillimore and K. Dexter for providing advice during the conceptualization of the study; L. Antão and M. Vellend for providing feedback on the draft manuscript; and the anonymous reviewers whose comments greatly enhanced our work. **Funding:** The BioTIME database was supported by ERC AdG BioTIME 250189 and ERC PoC BioCHANGE 727440. We thank the ERC projects BioTIME and BioCHANGE for supporting the initial data synthesis work that led to this study, and the Leverhulme Centre for Anthropocene Biodiversity for continued funding of the database. Also supported by a Carnegie-Caledonian PhD Scholarship and NERC doctoral training partnership grant NE/L002558/1 (G.N.D.), a Leverhulme Fellowship and the Leverhulme Centre for Anthropocene Biodiversity (M.D.), Leverhulme Project Grant RPG-2019-402 (A.E.M. and M.D.), and the German Centre of Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (funded by the German Research Foundation; FZT 118, S.A.B.). **Author contributions:** G.N.D., M.D., and I.H.M.-S. conceptualized the study; G.N.D. integrated databases and conducted statistical analyses with input from S.A.B., I.H.M.-S., A.D.B., and M.D.; G.N.D. created the figures with input from co-authors; S.A.B., M.D., and S.R.S. wrote the code for the rarefaction of the BioTIME studies; G.N.D. wrote the first draft; and all authors contributed to revisions. I.H.M.-S. was the primary supervisor, M.D. the co-supervisor, and A.D.B. is on the supervisory committee for G.N.D. A.E.M. and M.D. fund the compilation of the BioTIME database. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Code for the rarefaction of the BioTIME Database is available at https://doi.org/10.5281/zenodo.1475218. Code for statistical analyses is available at http://doi.org/10.5281/zenodo.1490144. Population and biodiversity data are freely available in the Living Planet and BioTIME databases (*25*, *26*). The Living Planet Database can be accessed on www.livingplanetindex.org/data_portal. The BioTIME Database can be accessed on Zenodo at https://doi.org/10.5281/zenodo.1211105 or through the BioTIME website at http://biotime.st-andrews.ac.uk. The public studies that were included in the version of BioTIME we analyzed can be downloaded from http://biotime.st-andrews.ac.uk/BioTIME_download.php. Land-use change data are publicly available in the Land Use Harmonization Database (*22*), the Forest Cover Change Database (*23*), and the MODIS Landcover Database (*24*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1341/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 to S4
References (*56*–*271*)

6 November 2019; accepted 7 April 2020
10.1126/science.aba1289

REPORT

METALLURGY

Making ultrastrong steel tough by grain-boundary delamination

L. Liu[1]*, Qin Yu[2]*, Z. Wang[1], Jon Ell[2,3], M. X. Huang[1]†, Robert O. Ritchie[2,3]†

Developing ultrahigh-strength steels that are ductile, fracture resistant, and cost effective would be attractive for a variety of structural applications. We show that improved fracture resistance in a steel with an ultrahigh yield strength of nearly 2 gigapascals can be achieved by activating delamination toughening coupled with transformation-induced plasticity. Delamination toughening associated with intensive but controlled cracking at manganese-enriched prior-austenite grain boundaries normal to the primary fracture surface dramatically improves the overall fracture resistance. As a result, fracture under plane-strain conditions is automatically transformed into a series of fracture processes in "parallel" plane-stress conditions through the thickness. The present "high-strength induced multidelamination" strategy offers a different pathway to develop engineering materials with ultrahigh strength and superior toughness at economical materials cost.

Lightweight yet cost-efficient structural materials with superior mechanical properties are continually sought after in engineering applications to meet the demands of sustainable economy development. Unfortunately, attaining high strength is usually at the expense of deteriorating toughness, which invariably is a major concern for safety-critical applications (*1*, *2*). The strengthening mechanisms in structural metals and alloys are built on the fundamental principle of inhibiting, or blocking, dislocation slip through the introduction of various obstacles at different length scales (*3*, *4*). However, abundant dislocations entangled in small imperfections can give rise to the localized stress concentrations that can cause crack initiation, which eventually can lead to catastropic failure (*1*, *5*). A mainstream effort to overcome the strength-toughness trade-off is focused on tailoring the microstructure or designing materials by means of solid-solution alloying. Multielement high- and medium-entropy alloys possess exceptional damage tolerance at cryogenic temperatures because of a deformation-induced nano-twinning mechanism (*6*, *7*). Maraging steels are another example and are recognized as the strongest metal alloys with acceptable damage tolerance for aerospace applications. However, maraging steels contain a large amount of costly alloying elements such as nickel (17 to 19 wt %), cobalt (8 to 12 wt %), and molybdenum (3 to 5 wt %) (*8*). Although the maraging steel alloying strategy is a perfect vehicle to attain superior mechanical performance, economical mass production and recycling are not feasible because of costs and environmental concerns (*3*, *9*). Reducing the average grain size is a different nonalloying pathway to enhance the strength-toughness combination. However, this approach can be limiting because the strengthening is usually achieved at the sacrifice of losing ductility (*10*, *11*). We show that exceptional damage tolerance can be achieved in an ultrastrong steel, with a simple composition and cost-effective processing route for fabrication. We demonstrate that increasing the yield strength is not detrimental to the toughness but instead can facilitate the activation of a delamination toughening (*12*, *13*) mechanism. This substantially enhances the toughness. Specifically, the ultrahigh-yield strength enables a secondary fracture mode, delamination cracking, at interfaces normal to the primary fracture surface. Multiple separated laminated ligaments develop near the fracture surface because of the delamination events, providing an extra energy release rate for fracture as well as enhancing crack-tip blunting, collectively increasing the overall fracture toughness. Such delamination toughening combined with transformation-induced plasticity (TRIP) toughening are rarely realized simultaneously in structural materals. The combination enables an intriguing combination of strengh, ductility, and toughness properties in our steel.

Our steel has a chemical composition of Fe-9.95% Mn-0.44% C-1.87% Al-0.67% V (weight percent) and was fabricated by means of a deformed and partitioned treatment (fig. S1) (*14*). Starting with an almost fully austenitic microstructure (fig. S2), the prior-austenite grains were extensively elongated along the rolling direction (RD) during the initial hot rolling and warm rolling processes (*14*, *15*). The austenite partially transforms to martensite during the subsequent cold rolling, resulting in a lamellar martensite-austenite duplex microstructure (Fig. 1A), which contains a volume fraction of austenite, measured with neutron diffraction at 47.5% (fig. S3). Last, partitioning was performed to optmize the mechanical stability of retained austenite by means of C partitioning from martensite to austenite (fig. S4) (*14*). The martensitic matrix (α') is composed of nano-sized grains decorated with intensive dislocations (fig. S5). The dislocation density of martensite matrix was determined by means of neutron diffraction to be 2.43×10^{16} m^{-2}, which is at least one order higher than that of other martensitic steels (fig. S6); further details of this calculation based on neutron diffraction measurements are given in (*16*). The austenite (γ) phase displays a stretched lamellar-shaped grain structure with dimensions of hundreds of micrometers along the RD, dozens of micrometers in the transverse direction (TD), and a few micrometers in the normal direction (ND) (Fig. 1A). The prior-austenite grain boundaries (PAGBs) are retained during cold rolling, where some austenite grains transform to martensite; this is indicated with the black dotted lines in the Fig. 1B three-dimensional (3D) stereographic microstructure, which was reconstructed from the PAGB maps. Further atomic-scale composition analysis of the present steel by means of 3D atom probe topography (APT) shows segregation of Mn and C to the PAGB (Fig. 1C) (*16*). No segregation of P, S, or other harmful elements to the PAGB was detected. To better illustrate the laminated duplex microstructure, we constructed a schematic 3D model (Fig. 1D).

To evaluate the mechanical properties of the ultrastrong steel, we characterized the tensile properties and the *J*-integral–based crack-resistance (J-Δa) R-curves in both the RD and TD orientations in ambient air. The RD and TD dog-bone–shaped tensile specimens were strained along and perpendicular to, respectively, the elongated austenitic grains (Fig. 2A). Correspondingly, the notch and crack propagation directions in the RD and TD compact-tension [C(T)] fracture toughness specimens were aligned perpendicular and parallel, respectively, to the elongated austenitic grains. Benefiting from the unusually high dislocation density, tensile loading along the RD yielded a superior combination of strength and ductility (Fig. 2B) (*14*). Specifically, the upper yield strength (σ_{yu}), the ultimate tensile strength (σ_{uts}), and the uniform elongation (ε_u) were determined to be 1978 MPa, 2144 MPa, and 19.0%, respectively (Fig. 2B and table S1). Properties in the loading direction along the TD are also sound, with a very high ultimate tensile strength of 2048 MPa, which is similar to that for the

[1]Department of Mechanical Engineering, University of Hong Kong, Pokfulam Road, Hong Kong, China. [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, CA 94720, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: mxhuang@hku.hk (M.X.H.); roritchie@lbl.gov (R.O.R.)

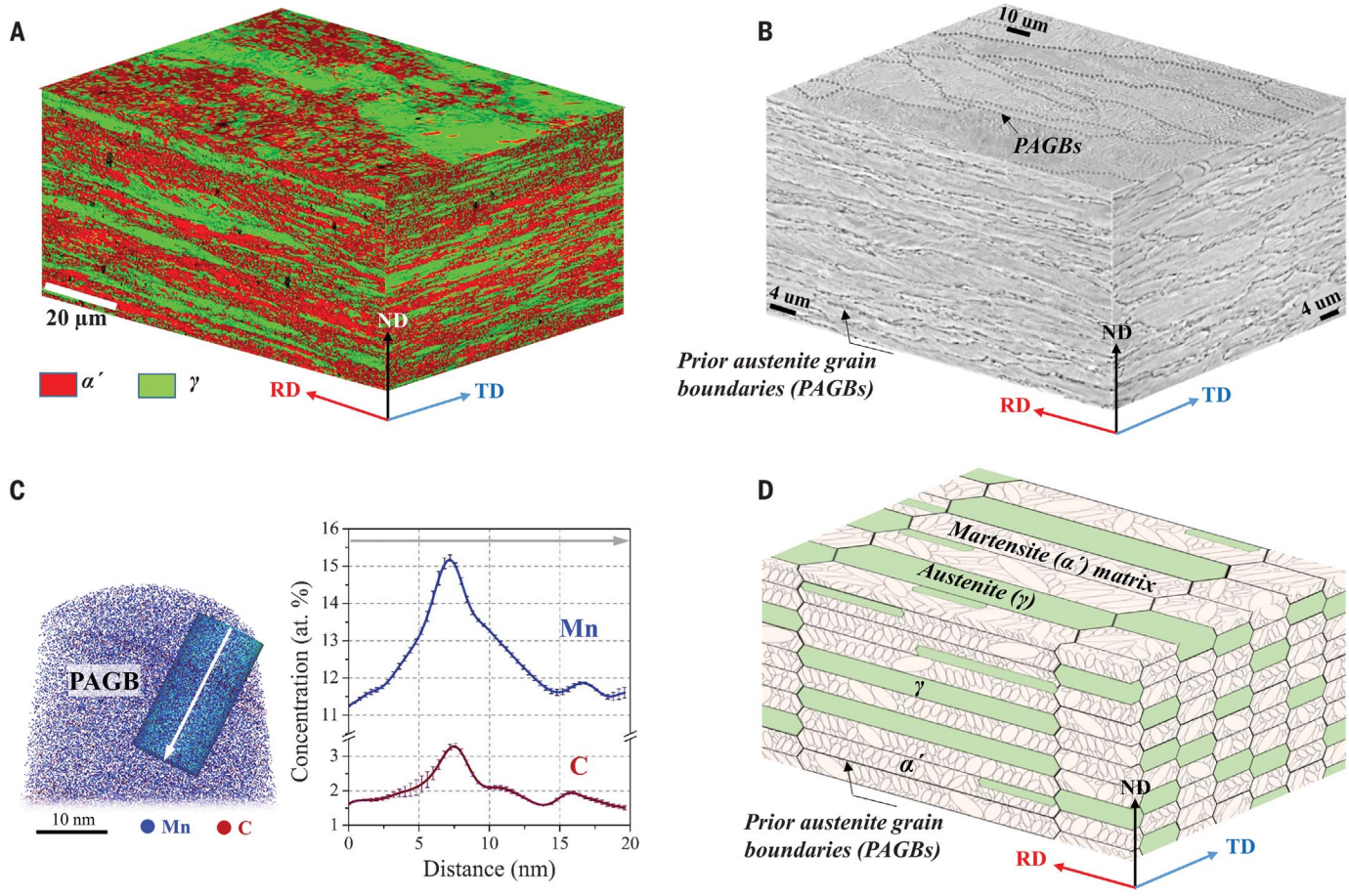


Fig. 1. Microstructure of the present steel. (**A**) 3D stereographic microstructure reconstructed with electron backscatter diffraction (EBSD) phase maps scanned on the orthotropic planes perpendicular to the RD, TD, and ND, respectively. A heterogeneous laminated duplex microstructure comprising martensitic matrix (α′) and elongated austenite (γ) lamellae is exhibited in the present steel. Retained austenite grains are elongated along the RD and slightly stretched along the TD. (**B**) 3D stereographic microstructure reconstructed by using the PAGBs maps imaged with SEM. The PAGBs are indicated by the black dotted lines and show a heterogeneous distribution. (**C**) 3D ion concentration map, which contains a PAGB and a corresponding 1D concentration profile across the boundary and shows the segregation of Mn and C to the PAGB. (**D**) Schematic 3D model illustrates the microstructural features of the heterogeneous microstructure of the present steel.

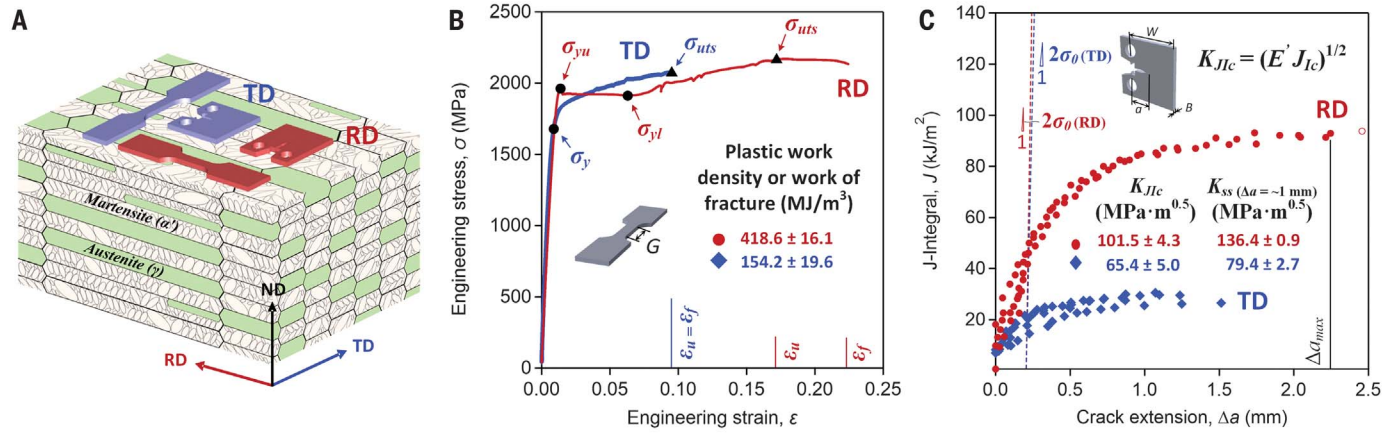


Fig. 2. Tensile and fracture properties of the present steel. (**A**) Schematic diagram describing the orientations of the dog bone–shaped tensile specimens and the C(T) specimens relative to the thin-sheet steel. (**B**) Engineering stress-strain curves of the present steel deformed under tension along the RD and TD orientations. (**C**) The *J*-integral–based resistance curves (*J*-R curves) measured from the C(T) specimens at room temperature.

RD orientation, but plastic deformation begins to proceed earlier; the 0.2%-offset yield strength (σ_y) in the TD orientation was measured to be 1714 MPa.

In view of its ultrahigh strength and plastic deformation capacity, we evaluated the fracture resistance of the present steel by measuring *J*-integral–based R-curves—*J* as a function of the stable crack extension, Δa—using the C(T) specimens in accordance with the ASTM Standard E1820 (*17*). Our steel displays a modest fracture toughness when loaded along the TD, presenting an average crack-initiation toughness of J_{Ic} = 19.6 kJ·m^{-2} (Fig. 2C and table S1). The crack-resistance (R-curve) behavior in the TD slightly rises as the crack extends, resulting in a valid crack-growth toughness of J_{ss} = 28.7 kJ·m^{-2} at a crack extension Δa of ~1 mm. These fracture properties are comparable with those of the best strong-yet-tough materials, such as 18Ni 300-grade maraging steels (*18*, *19*) and commercial aircraft–quality 300M and 4340 steels (*20*, *21*). In spite of a higher yield strength of the RD tensile specimens, the R-curve in the RD orientation reveals an even better crack resistance, showing a J_{Ic} of 46.9 kJ·m^{-2} and a crack-growth toughness J_{ss} of 84.6 kJ·m^{-2} as the crack extends to Δa ~ 1 mm (Fig. 2C and table S1). These toughness values in the RD are almost 1.5 and 2 times higher than those in the TD, respectively. According to the standard mode-I *J-K* equivalence relationship, the crack-initiation toughness

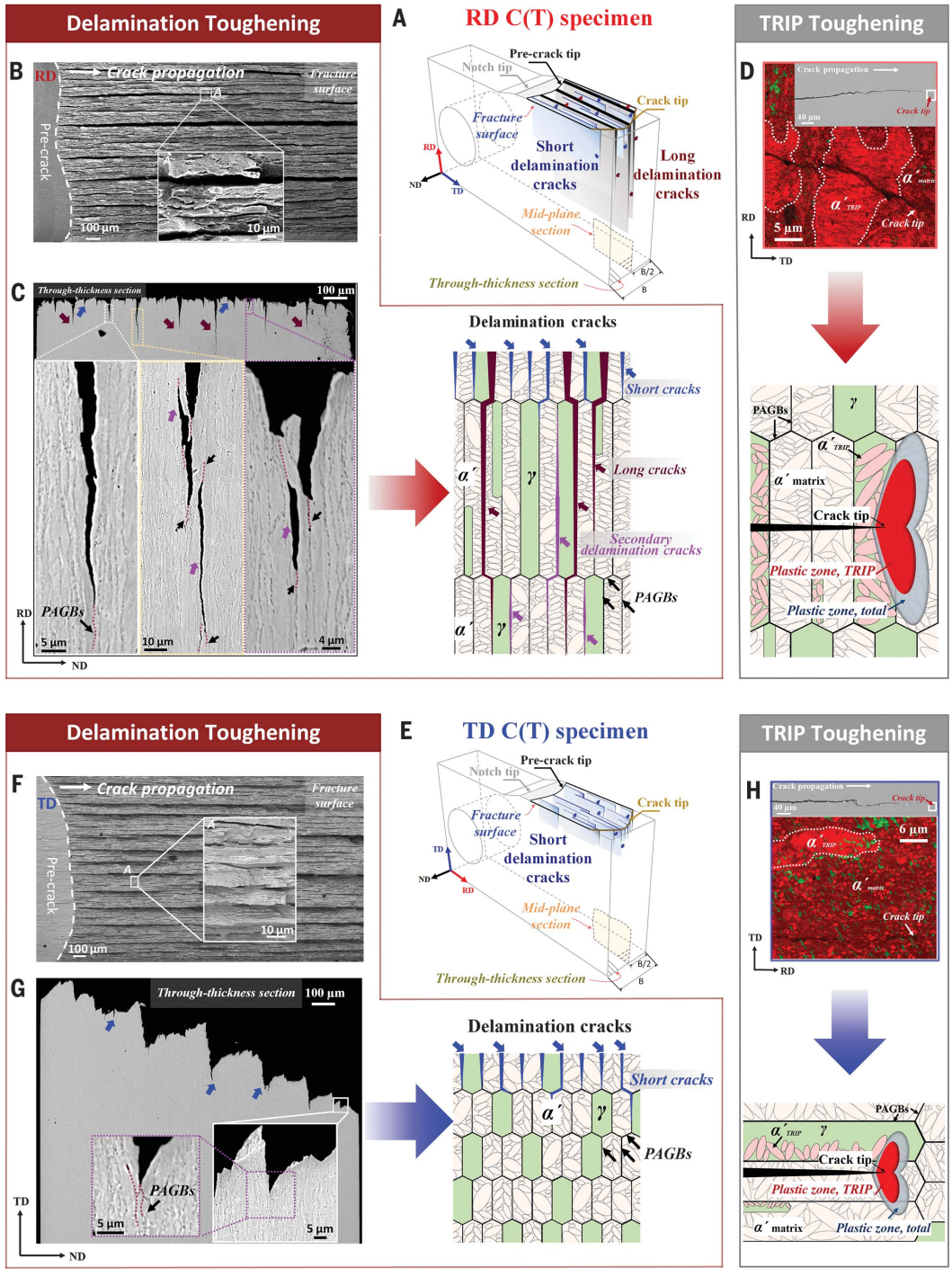


Fig. 3. Toughening mechanisms in the steel. (**A** and **E**) The schematic diagrams of the RD and TD C(T) specimens, respectively, showing the various sections for microstructure characterizations. (**B** and **F**) The fracture surfaces of the RD and TD C(T) specimens, clearly showing the thin-layer delamination bands on the fracture surface. (**C** and **G**) The SEM images captured on the through-thickness section normal to the fracture surface show the development of delamination cracks along PAGBs. Short delamination cracks (blue arrows) and deeply penetrated ones (dark red arrows) are found in the RD specimen, whereas only shallow-penetrated delamination cracks formed in the TD specimen. Slender cracks disconnected from the fracture surface (purple arrows) were usually observed in the vicinity of the long delamination cracks. Corresponding schematic diagrams reveal that the substantially elongated PAGBs in the RD specimen facilitate the propagation of delamination cracks, whereas the short PAGBs and numerous PAGBs perpendicular to the crack propagation in the TD specimen hinder the extension of delamination cracks. (**D** and **H**) EBSD phase maps overlaid with the image quality maps scanned in the vicinity of the main crack tip on the mid-plane section of C(T) specimens. Schematic diagrams delineating the TRIP-toughening mechanism are shown below. The TRIP-induced martensite (α'_{TRIP}) can be distinguished from the martensite matrix (α'_{matrix}) through higher image quality because the α'_{TRIP} grains transformed from the retained austenite have a lower dislocation density (regions bounded by dashed white lines).

(K_{JIc}) of the RD and TD specimens were determined to be 101.5 and 65.4 MPa·m^{½}, respectively. Likewise, the crack-growth toughness at Δa ~1 mm, K_{ss}, was 136.4 and 79.4 MPa·m^{½} for the RD and TD specimens, respectively. These are very high values of the crack-initiation and crack-growth toughnesses in our ultrastrong steel that are not found in any other existing structural materials at a comparable yield strength (~2 GPa).

To illuminate the underlying toughening mechanisms responsible for the exceptional damage tolerance of the ultrastrong steel, we characterized the microstructures on various sections of the RD and TD C(T) specimens (Fig. 3, A and E). Featured regions on the fracture surfaces characterized with scanning electron microscopy (SEM) show the existence of multiple thin-layer delamination bands in both RD and TD orientations (Fig. 3, B and F). Observed from the fracture surface, the original ~1.4-mm-thick RD specimens were delaminated through the thickness into several thin layers with their interspacings (or delamination cracks) "penetrating" into the material along the planes perpendicular to the fracture surface. Moreover, delamination cracks at different length scales were developed in the present steel, resulting in numerous delamination bands sized in the range of several micrometers (Fig. 3B, region A). The thickness of these delamination ligaments was substantially thinner than that in other structural materials that contain delamination cracks (*12*, *13*). Delamination was also present in the TD specimens, but the delamination cracks were fewer in number and shorter in length than those in the RD specimens. The microstructure in the vicinity of the delamination cracks in the through-thickness sections normal to the fracture surface was further characterized to clarify the micromechanisms associated with the delamination cracks (Fig. 3, C and G). On the basis of the statistical distribution of the crack lengths (figs. S7 and S8), the delamination cracks in the RD specimens can be categorized into two groups: short cracks with lengths <~50 μm (Fig. 3C, blue arrows) and longer cracks with lengths >~50 μm (Fig. 3C, dark red arrows). Moreover, slender cracks disconnected from the fracture surface (Fig. 3C, purple arrows) were usually observed in the vicinity of the long delamination cracks. Further characterization at the tips of the delamination cracks clearly indicated that these cracks propagate predominantly along the PAGBs (Fig. 3C). In the case of the TD specimens, only short delamination cracks with lengths shorter than 50 μm were observed (Fig. 3G). Clearly, loading along the RD results in a much larger tendency for delamination.

The activation of delamination toughening requires two necessary conditions: intrinsically and microstructurally, the existence of "relatively weak interfaces," where delamination takes place; and mechanically, with a "high mechanical stress" that exceeds the critical fracture stress of the "relatively weak interfaces." From our characterization of the locations at delamination crack tips, the Mn-enriched PAGBs in our steel serve as the relatively weak interfaces and preferential sites for the initiation and propagation of the delamination microcracks (Fig. 3, C and G). Actually, the Mn-enriched PAGBs still have a high level of cohesion strength but are just not as strong as the grain interior owing to the segregation of Mn (Fig. 1C) (*22*, *23*). By comparison, such delaminations do not occur in maraging steels, no matter how high their yield strengths are, because there are no such interfaces and preferential sites along the boundaries with reduced cohesion.

From the perspective of the mechanical driving force, a triaxial tensile stress-state exists ahead of a crack tip under plane-strain conditions. The material in front of the crack tip is subjected to a tensile stress (σ_2) (fig. S9A) along the thickness direction (Fig. 1A, ND). When the tensile stress σ_2 that is perpendicular to the PAGBs is sufficiently high to reach the critical fracture stress of the Mn-enriched PAGBs (Fig. 3A), delamination will occur. To meet this requirement, the material should have sufficiently high yield strength so that a high external applied stress is required. A higher external applied stress will lead to a higher σ_2 at the crack tip. For certain structural materials that contain weak interfaces, delamination cracks are mainly formed at low temperatures at which the interfaces may be more brittle and the through-thickness stresses are larger. This leads to the fascinating effect that such alloys—for example, certain Al–Li alloys (*12*) and ultrafine-grained low alloy steels (*24*, *25*)—become tougher at cryogenic temperatures compared with ambient temperature. As demonstrated here, the present steel possesses ultrahigh yield strength because of its high dislocation density and Mn-enriched PAGBs—with relatively low cohesion as compared with the grain interior—acting as the preferential delamination sites; this meets the two necessary requirements to activate delamination toughening and leads to its exceptional room-temperature fracture toughness for a metallic alloy with a yield strength in the 2 GPa range.

As a consequence of the multiple delamination cracks that occur ahead of the crack front and perpendicular to the crack plane, the expected fracture under plane-strain conditions is transformed into a series of fracture processes in "parallel" plane-stress ligaments through the thickness that individually display a far higher (plane-stress) toughness than for a single (full-thickness) plane-strain section (fig. S9B). In parallel, numerous new interfaces generated during the delamination process consume energy, which effectively increases the energy release rate to contribute to the exceptional fracture toughness. Our current study further reveals that the delamination toughening is affected by the relative orientations of the delamination crack path with respect to the

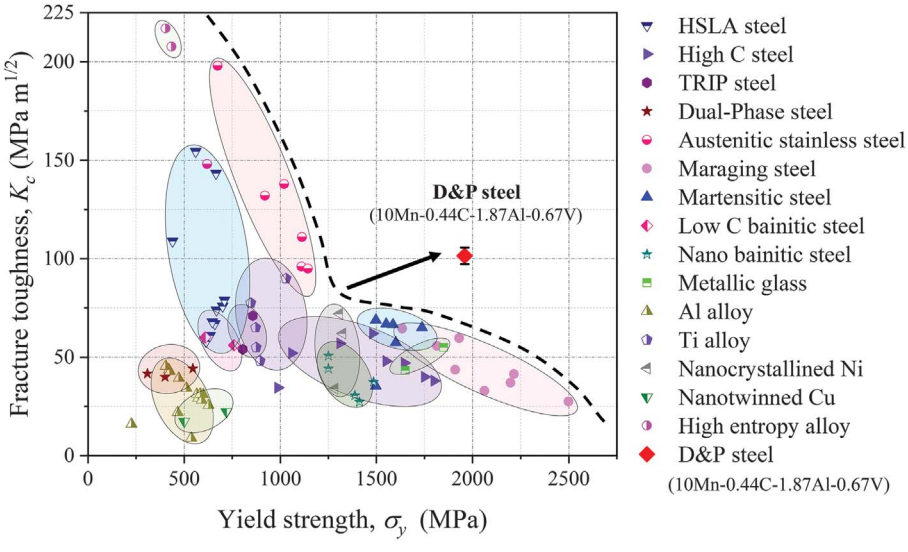


Fig. 4. Ashby map in terms of the fracture toughness versus the yield strength. Our ultrastrong steel overcomes the strength-toughness trade-off shown in most existing structural materials, especially high-strength low-alloy (HSLA) steels (*29*, *30*), high carbon (C) steels (*31*), TRIP steels (*32*), dual-phase steels (*33*), austenitic stainless steels (*34*, *35*), maraging steels (*18*, *19*, *36*), martensitic steels (*37*, *38*), low C bainitic steels (*39*), nano bainitic steels (*40*), metallic glass (*41*), Al alloys (*42*), Ti alloys (*43*), nanocrystallined Ni (*44*), nanotwinned Cu (*45*), and high-entropy alloys (*6*, *7*). Detailed composition and properties of these compared materials can be found in table S2.

elongated duplex structure. The lengths of the PAGBs along the RD are almost three times the lengths of the PAGBs parallel to the TD because of the large rolling reduction (Fig. 1B). As the delamination cracks propagate along the elongated PAGBs parallel to RD in the RD specimens, longer delamination cracks are developed (Fig. 3C). However, delamination cracks extend with short PAGBs parallel to the TD on the through-thickness section in the TD specimen. Furthermore, numerous PAGBs, aligned perpendicular to the crack path, are developed in the TD specimen. These grain boundaries are effective obstacles to retard the propagation of delamination cracks (Fig. 3G). Therefore, the longer delamination cracks developed in the RD specimens and the shorter delamination cracks developed in the TD specimens give rise to a larger and smaller toughening effect, respectively, which is consistent with the higher and lower respective values of the measured toughness in these orientations.

The present steel also displays a TRIP effect during crack advance. The TRIP-toughening is due to an in situ austenite-to-martensite transformation that is triggered in the vicinity of the crack tip, where an initial duplex martensitic-austenitic microstructure (Fig. 1A) is transformed to an almost fully martensitic microstructure (Fig. 3, D and H) after crack propagation. Because of the lattice-parameter difference between martensite and parent austenite, a compressive residual stress is likely introduced by the TRIP effect in the vicinity of crack tip to further resist crack propagation (*26*). In addition, TRIP effect can serve to promote strain hardening, which in turn enhances the uniform ductility by delaying the onset of the necking instability to provide additional intrinsic toughening to resist crack propagation (*27*, *28*).

Similar to the delamination toughening, the TRIP-toughening is also a function of the crack extension orientations. A larger-sized TRIP zone is realized in the RD specimen (Fig. 3D), where the main crack propagates perpendicular to the elongated austenite. By contrast, a small-sized TRIP zone is produced (Fig. 3H) in the TD specimen because the growth of the TRIP zone is substantially restricted by the PAGBs parallel to the elongated austenite lamellae (*28*). This is illustrated in the schematic figures in Fig. 3, D and H. Consistently, higher toughness values exhibited in the RD specimens are ascribed to the stronger TRIP toughening effect induced by the larger-sized TRIP zone, and lower toughness values exhibited in the TD specimens are ascribed to the weaker TRIP toughening effect induced by the smaller-sized TRIP zone.

To demonstrate how the damage tolerance in our steel compares with other ultrahigh-strength structural materials, we show in Fig. 4 an Ashby map of the crack-initiation fracture toughness versus yield strength. Our steel exhibits a yield strength comparable with some of the strongest existing metallic materials—namely, maraging steels—but with an initiation toughness (K_c) that is almost a factor of two greater. Our steel displays a toughness comparable with that of titanium alloys but with a strength that is greater by a factor of two. The combination of high strength and high toughness clearly demonstrates that the "high-strength induced multidelamination" mechanism can be highly effective in maximizing the mechanical properties of high-strength structural materials while minimizing material cost. Our material design principle, which exploits high strength combined with relatively weak interfaces, is one that we believe can be widely applied to optimize the mechanical performance of materials with ultrahigh strength.

REFERENCES AND NOTES

1. R. O. Ritchie, *Nat. Mater.* **10**, 817–822 (2011).
2. K. Lu, *Science* **328**, 319–320 (2010).
3. X. Li, K. Lu, *Nat. Mater.* **16**, 700–701 (2017).
4. K. Lu, L. Lu, S. Suresh, *Science* **324**, 349–352 (2009).
5. M. JW, *ISIJ Int.* **41**, 599–611 (2001).
6. B. Gludovatz *et al*., *Science* **345**, 1153–1158 (2014).
7. B. Gludovatz *et al*., *Nat. Commun.* **7**, 10602 (2016).
8. S. Floreen, *Metallurg. Rev.* **13**, 115–128 (1968).
9. X. Li, K. Lu, *Science* **364**, 733–734 (2019).
10. K. Lu, *Science* **345**, 1455–1456 (2014).
11. Y. Wang, M. Chen, F. Zhou, E. Ma, *Nature* **419**, 912–915 (2002).
12. K. T. Venkateswara Rao, W. Yu, R. O. Ritchie, *Metall. Trans., A, Phys. Metall. Mater. Sci.* **20**, 485–497 (1989).
13. K. S. Chan, *Metall. Trans., A, Phys. Metall. Mater. Sci.* **20**, 155–164 (1989).
14. B. B. He *et al*., *Science* **357**, 1029–1032 (2017).
15. L. Liu, B. B. He, M. X. Huang, *Metals (Basel)* **9**, 695 (2019).
16. Materials and methods are available as supplementary materials.
17. ASTM International, *E1820-17a Standard Test Method for Measurement of Fracture Toughness* (ASTM International, 2017).
18. C. Carter, *Eng. Fract. Mech.* **3**, 1–13 (1971).
19. Y. He, K. Yang, W. Qu, F. Kong, G. Su, *Mater. Lett.* **56**, 763–769 (2002).
20. R. O. Ritchie, *Met. Sci.* **11**, 368–381 (1977).
21. R. O. Ritchie, B. Francis, W. L. Server, *Metall. Trans., A, Phys. Metall. Mater. Sci.* **7**, 831–838 (1976).
22. M. Kuzmina, D. Ponge, D. Raabe, *Acta Mater.* **86**, 182–192 (2015).
23. N. Heo, J. Nam, Y.-U. Heo, S.-J. Kim, *Acta Mater.* **61**, 4022–4034 (2013).
24. Y. Kimura, T. Inoue, F. Yin, K. Tsuzaki, *Science* **320**, 1057–1060 (2008).
25. J. W. Morris Jr., *Science* **320**, 1022–1023 (2008).
26. W. Bleck, X. Guo, Y. Ma, *Steel Res. Int.* **88**, 1700218 (2017).
27. R. H. Leal, Ph.D. Thesis, Massachusetts Institute of Technology (1984).
28. M. M. Wang, C. C. Tasan, D. Ponge, A. C. Dippel, D. Raabe, *Acta Mater.* **85**, 216–228 (2015).
29. M. Linaza, J. Romero, J. Rodriguez-Ibabe, J. Urcola, *Scr. Metall. Mater.* **29**, 1217–1222 (1993).
30. M. E. Haque, K. Sudhakar, *Int. J. Fatigue* **24**, 1003–1010 (2002).
31. S. K. Putatunda, *Mater. Sci. Eng. A* **297**, 31–43 (2001).
32. J. Kobayashi, D. Ina, A. Futamura, K.-i. Sugimoto, *ISIJ Int.* **54**, 955–962 (2014).
33. A. Bayram, A. Uğuz, M. Ula, *Mater. Charact.* **43**, 259–269 (1999).
34. L. Xiong, Z. You, S. Qu, L. Lu, *Acta Mater.* **150**, 130–138 (2018).
35. J. E. Pawel *et al*., *J. Nucl. Mater.* **212-215**, 442–447 (1994).
36. L. Van Swam, R. Pelloux, N. Grant, *Powder Metall.* **17**, 33–45 (1974).
37. R. O. Ritchie, *J. Eng. Mater. Technol.* **99**, 195–204 (1977).
38. G. Lai, W. Wood, R. Clark, V. Zackay, E. R. Parker, *Metall. Trans.* **5**, 1663–1670 (1974).
39. H. Aglan, Z. Liu, M. Hassan, M. Fateh, *J. Mater. Process. Technol.* **151**, 268–274 (2004).
40. C. Garcia-Mateo, F. Caballero, *ISIJ Int.* **45**, 1736–1740 (2005).
41. M. D. Demetriou *et al*., *Nat. Mater.* **10**, 123–128 (2011).
42. N. Kamp, I. Sinclair, M. Starink, *Metall. Mater. Trans., A Phys. Metall. Mater. Sci.* **33**, 1125–1136 (2002).
43. M. Niinomi, *Mater. Sci. Eng. A* **243**, 231–236 (1998).
44. R. Mirshams, C. Xiao, S. Whang, W. Yin, *Mater. Sci. Eng. A* **315**, 21–27 (2001).
45. A. Singh, L. Tang, M. Dao, L. Lu, S. Suresh, *Acta Mater.* **59**, 2437–2446 (2011).

ACKNOWLEDGMENTS

L. H. He and J. Chen are acknowledged for their help on the neutron powder diffraction experiments which were performed at the general purpose powder diffractometer of the China Spallation Neutron Source (CSNS), Dongguan, China. J. H. Luan and Z. B. Jiao are acknowledged for their help on the APT experiments conducted at the Inter-University 3D Atom Probe Tomography Unit of City University of Hong Kong. We thank B. B. He for his assistance with the preparation of the material and M. Wang for help with the determination of the dislocation density. Last, we thank K. Lu for his insightful comments on the paper. **Funding:** M.X.H. acknowledges the financial support from the National Key Research and Development Program of China (2019YFA0209900 and 2017YFB0304401), National Natural Science Foundation of China (U1764252), and Research Grants Council of Hong Kong (R7066-18, 17255016, and 17210418). Q.Y. and R.O.R. acknowledge financial support for the Mechanical Behavior of Materials Program (KC13) at the Lawrence Berkeley National Laboratory (LBNL) provided by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division under contract DE-AC02-05-CH11231. EBSD experiments were carried out at LBNL's Molecular Foundry supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under the same contract number. **Author contributions:** M.X.H., Q.Y., and R.O.R. designed the research. M.X.H. and R.O.R. supervised the study. L.L. fabricated the steel material. Q.Y., J.E., and L.L. performed the mechanical characterization. L.L., Q.Y., and Z.W. conducted the SEM and EBSD characterizations. L.L. worked on the transmission electron microscopy characterization. L.L. and Z.W. conducted the APT measurement and the neutron diffraction characterization. L.L., Q.Y., M.X.H., and R.O.R. analyzed the data and wrote the manuscript. All the authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data are available in the manuscript and supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1347/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 and S2
References (*46*–*54*)

17 January 2020; accepted 27 April 2020
10.1126/science.aba9413

ELECTROCHEMISTRY

Microfluidic electrochemistry for single-electron transfer redox-neutral reactions

Yiming Mo[1]*, Zhaohong Lu[2]*, Girish Rughoobur[3], Prashant Patil[4], Neil Gershenfeld[4], Akintunde I. Akinwande[3], Stephen L. Buchwald[2]†, Klavs F. Jensen[1]†

Electrochemistry offers opportunities to promote single-electron transfer (SET) redox-neutral chemistries similar to those recently discovered using visible-light photocatalysis but without the use of an expensive photocatalyst. Herein, we introduce a microfluidic redox-neutral electrochemistry (μRN-eChem) platform that has broad applicability to SET chemistry, including radical-radical cross-coupling, Minisci-type reactions, and nickel-catalyzed C(sp^2)–O cross-coupling. The cathode and anode simultaneously generate the corresponding reactive intermediates, and selective transformation is facilitated by the rapid molecular diffusion across a microfluidic channel that outpaces the decomposition of the intermediates. μRN-eChem was shown to enable a two-step gram-scale electrosynthesis of a nematic liquid crystal compound, demonstrating its practicality.

Over the past decade, pioneering developments of visible-light photocatalysis in organic synthesis have enabled previously inaccessible redox-neutral reactions that proceed through single-electron transfer (SET) processes (*1*, *2*). Nonetheless, the use of photocatalysts, mostly precious metal complexes (*3*) or sophisticated organic dyes (*4*), could have practical limitations, such as the nontrivial tuning of redox potentials; high cost of transition-metal photocatalysts at scale (*5*); incompatibility of photocatalysts with strong nucleophiles, electrophiles, or radical intermediates (*6*, *7*); and challenging removal of transition metals during purification of the products (*8*, *9*). Electrosynthesis, on the other hand, is an emerging redox platform accessing environmentally benign, cost-effective, scalable, and distinctive transformations (*10*) powered by inexpensive electricity. Consequently, we considered whether electrochemistry could be applied in a practical photocatalyst-free system for SET redox-neutral reactions.

Most of the reported synthetic electrochemistry relies on reactions on a single electrode with by-products generated on the other electrode, and, as such, the nature of the desired electrochemical transformations is either oxidative or reductive (*10*). In contrast, redox-neutral electrochemistry (i.e., paired or coupled electrosynthesis), which involves two desirable half-electrode reactions performed simultaneously, is underdeveloped, despite its relative material and energy efficiency (*10*–*12*), and examples of radical-based redox-neutral electrosynthesis are even rarer (*13*–*15*). In conventional paired electrochemistry setups, the difficulties associated with matching the generation and interelectrode transport rates of the different highly reactive intermediates pose substantial obstacles to achieving the selective transformation over alternative undesired pathways. Microfluidics, however, has been shown to offer controllable and rapid species transport within a micrometer channel (*16*), with recent applications in electrosynthesis for improved reaction performance (*17*, *18*). To further exploit the capability of microfluidic electrochemistry, we sought to develop a microfluidic redox-neutral electrochemical (μRN-eChem) platform to overcome the challenges of photochemistry-inspired SET redox-neutral reactions that involve reactive intermediates generated from both electrodes.

SET radical-radical cross-coupling reactions were selected as the target transformations, because they offer a strategy to perform C(sp^2)–C(sp^3) bond formation without transition-metal catalysts (*19*). In photocatalytic implementations, a persistent radical and a transient radical are consecutively generated within a photocatalytic cycle in homogeneous solution and subsequently combine to give the selective cross-coupling product (*20*, *21*) (Fig. 1A). In contrast, in the electrochemical system (Fig. 1B), the delocalized generation of the persistent radical (P•) and the transient radical (T•) from two separate electrodes requires P• to migrate to the surface of the other electrode surface in order to react with T•. In conventional electrochemical setups, in which electrode distances typically range from millimeters to centimeters, the selective cross-coupling of P• and T• is hindered by radical decomposition due to the inefficient mixing of the two radicals. In a microfluidic parallel flow channel, the interelectrode radical transport is primarily controlled by molecular diffusion, the characteristic time of which can be estimated by $t = d^2/D$ (*22*), where d is the interelectrode distance, and D is the molecular diffusivity. Hence, we hypothesized that an extremely thin interelectrode gap flow cell would permit rapid diffusion that can outpace the radical decomposition, selectively yielding the radical-radical cross-coupling product.

To test the proposed μRN-eChem platform, we selected model precursors for the transient and persistent radicals (Fig. 1C). Decarboxylative Kolbe electrolysis (*23*), the oldest electrosynthetic reaction (discovered in 1847), exemplifies an anodic alkyl radical generation pathway that is attractive owing to the wide availability of carboxylic acids. For the cathodic persistent radical generation, we selected electron-deficient aryl nitriles (*24*) as precursors. An electron-deficient aryl nitrile **2** can undergo a SET event on the cathode to form a persistent radical anion **4** at a half-wave potential ($E_{1/2}$) of –2.01 V [versus the ferrocene/ferrocenium redox couple (Fc/Fc^+) on glassy carbon (GC) in acetonitrile (MeCN)]. Meanwhile, the deprotonation of alkyl carboxylic acid **1** with a Brønsted base gives an alkyl carboxylate anion, which can undergo SET oxidation at a half-peak potential ($E_{p/2}$) of +0.93 V (versus Fc/Fc^+ on GC in MeCN) with subsequent rapid loss of CO_2, yielding a transient alkyl radical **3**. Although such a broad electrochemical window would typically require screening and tuning of photocatalysts to achieve thermodynamically favorable radical generation (*25*), electrochemistry can easily initiate the redox reactions at the proper potential setting.

We engineered a μRN-eChem flow cell with a variable interelectrode distance (25 to 500 μm). Two GC plate electrodes, micromachined by a 532-nm laser, sandwich a thin fluorinated ethylene propylene film to create the microfluidic channel (see fig. S1). 4,4′-Biphenyldicarbonitrile (BPDN, **6**) was selected as a representative persistent radical precursor, and ultraviolet-visible (UV-Vis) spectroelectrochemical measurement of its radical anion **7** showed zero-order decomposition kinetics (Fig. 1D). This radical anion was tested in cross-coupling reactions with two types of transient C(sp^3)-centered radicals: a secondary alkyl radical with stabilization by an α-N and a primary alkyl radical without stabilization, generated from corresponding carboxylic acids **8** and **10**, respectively (Fig. 1E). The yields of both cross-coupling products (**9** and **11**) diminished with increasing interelectrode distance. Notably, the yield for cross-coupling of **7** with the α-N stabilized transient radical was less sensitive to interelectrode distance compared with the more reactive primary radical without stabilization. Furthermore, in a batch setup, **9** can still be produced in a reasonable yield (40%) (*15*), whereas only a very low yield of coupling product

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Center for Bits and Atoms, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
†Corresponding author. Email: kfjensen@mit.edu (K.F.J.); sbuchwal@mit.edu (S.L.B.) *These authors contributed equally to this work.

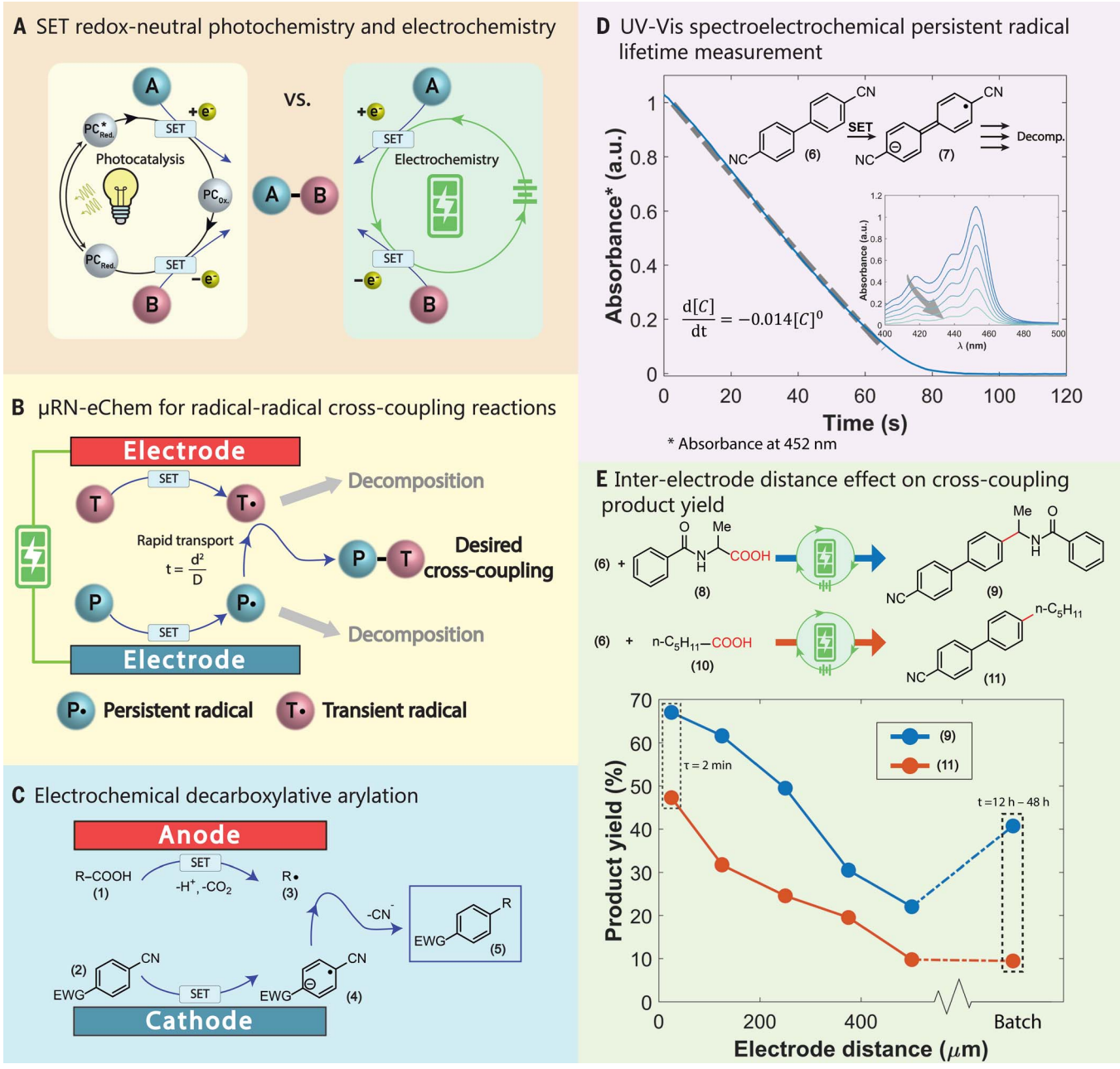


Fig. 1. Background and microfluidic redox-neutral electrochemistry (μRN-eChem). (**A**) Comparison of photochemistry and electrochemistry for SET redox-neutral reactions. PC, photocatalyst; PC_{Red}, ground-state PC; PC^*_{Red}, excited-state PC; PC_{Ox}, oxidized-state PC. (**B**) Concept of μRN-eChem for the cross-coupling reaction of persistent and transient radicals. (**C**) Mechanism of redox-neutral electrochemical cross-coupling reaction of carboxylic acids and electron-deficient aryl nitriles. R–COOH, carboxylic acid (where R is the alkyl group); R•, alkyl radical; EWG, electron-withdrawing group. (**D**) UV-Vis spectroelectrochemical lifetime measurement of BPDN radical anion. a.u., arbitrary units; Decomp., decomposition; [*C*], concentration; λ, wavelength. (**E**) Effect of interelectrode distance on reaction yield. Two batch setups were tested (fig. S5), and the setup with higher yield is presented in this plot. Me, methyl.

(9%) was obtained for **11**. The smallest interelectrode distance we examined (25 μm) offers a subsecond molecular diffusion time that is substantially shorter than the lifetime of **7**, leading to a greatly improved cross-coupling selectivity compared with conventional electrochemical setups.

Reaction optimization (see table S1) within the 25-μm flow cell identified Bu_4NOH and MeCN as the optimal base and solvent, respectively. Six equivalents of the carboxylic acids were used to obtain optimal yields; however, a smaller quantity (3 equiv) can be used, albeit with slightly diminished efficiency (9% reduction in yield). No additional supporting electrolyte was required, because of the excellent conductivity in the microfluidic channel. Next, we focused our attention on the scope of carboxylic acids and aryl nitriles that could be successfully transformed (Fig. 2). An extensive range of carboxylic acids with various functional groups proved to be suitable substrates in the cross-coupling with 1,4-dicyanobenzene (DCB) (**12** to **21**). Aliphatic carboxylic acids, which generate highly energetic $C(sp^3)$ alkyl radicals without stabilization that often lead to catalyst decomposition in photochemical

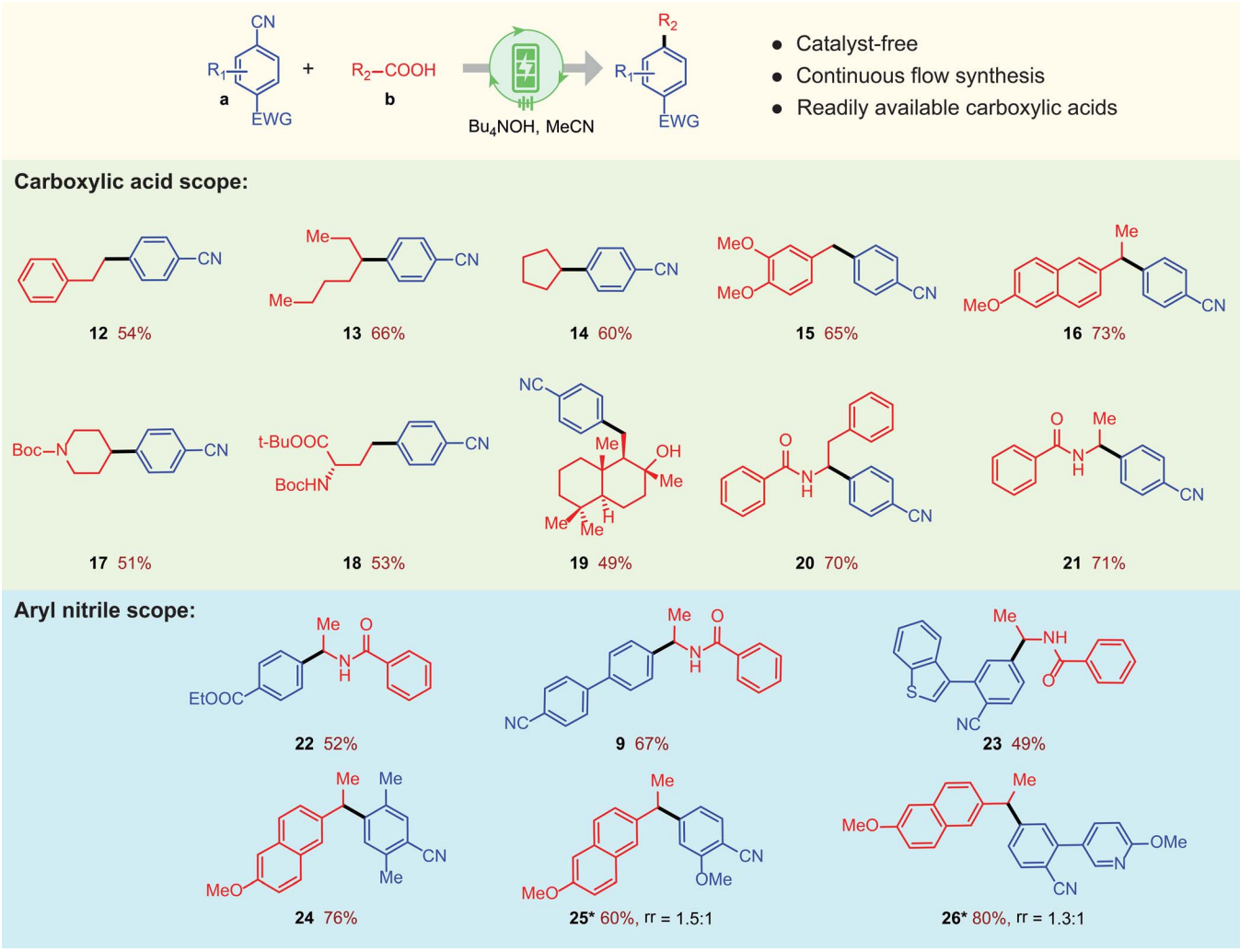


Fig. 2. Substrate scope of decarboxylative arylation in continuous-flow synthesis enabled by μRN-eChem. See the supplementary materials for detailed reaction conditions for each substrate; reactions were performed on a 0.4-mmol scale, unless otherwise noted. Asterisks indicate isomers observed, and in all cases the major isomer is depicted. Yields refer to the combined yield of all isomers. Boc, *tert*-butyloxycarbonyl; t-Bu, *tert*-butyl; Et, ethyl; rr, regioisomeric ratio.

systems (*6*, *7*), were well tolerated under our μRN-eChem conditions (**12** to **14** and **17** to **19**, 49 to 66% yield). Benzylic carboxylic acids, including the anti-inflammatory drug naproxen (**16b**), could be smoothly transformed to 1,1-diarylalkanes (**15** and **16**, 65 and 73%, respectively). Additionally, 77% of unreacted **16b** (based on the quantity of starting material used) was separated, demonstrating the recovery of excess carboxylic acid. Benzoyl-protected α-amino acids, delivering α-amino radicals, can also serve as good coupling partners to produce benzylic amides (**20** and **21**, 70 and 71% yield, respectively). The scope of electron-deficient aryl nitriles has also been examined. An ester group was a suitable alternative to the nitrile electron-withdrawing group (**22**, 52% yield). The BPDN radical anion ($BPDN^{\bullet-}$) **7**, despite its relatively reduced stability compared with the DCB radical anion ($DCB^{\bullet-}$) (Fig. 1D and fig. S3), coupled with α-amino radical in high efficiency (**9**, 67%). Derivatives of DCB, including those bearing heterocycle substituents, were effective for radical-radical cross-coupling to give corresponding coupling products (**23** to **26**, 49 to 80% yield).

Given the successful implementation of μRN-eChem with decarboxylative arylation reactions, we posited that other transient radical precursors could react in a similar manner. We first tested an electrochemical α-amino C–H arylation reaction (Fig. 3A) (*26*). Anodic oxidation of amine **27** [$E_{p/2}$ = +0.50 V (versus Fc/Fc^{+})] generates the radical cation **28**, and deprotonation of the methylene group α to the N gives α-aminoalkyl radical **29**, which undergoes cross-coupling with the cathodically generated persistent radical **31** to form arylation products (**33** and **34**, 69 and 64% yield, respectively). Electrochemical deboronation of trifluoroborate **35** [$E_{p/2}$ = +0.55 V (versus Fc/Fc^{+})] is a similar process to anodic decarboxylation, yielding a $C(sp^3)$ alkyl radical **36** (*27*, *28*) (Fig. 3B). Radical-radical coupling of **31** and **36** with subsequent decyanation gave **38** and **39** with good efficiency (73 and 50%, respectively). This μRN-eChem radical-radical cross-coupling strategy can also function synergistically with other catalytic mechanisms. We demonstrated the thiol-catalyzed allylic arylation reaction as an example (Fig. 3C). Electrophilic thiyl radical **40**, generated from anodic oxidation of thiol **41** under basic conditions [$E_{p/2}$ = +0.41 V (versus Fc/Fc^{+})], can regioselectively abstract an allylic hydrogen atom from alkene **42** to provide the allylic radical **43** (*29*). Rapid trapping of **43** using a persistent radical **31** forges a new $C(sp^2)$–$C(sp^3)$ bond (**45** and **46**, 83 and 65% yield, respectively).

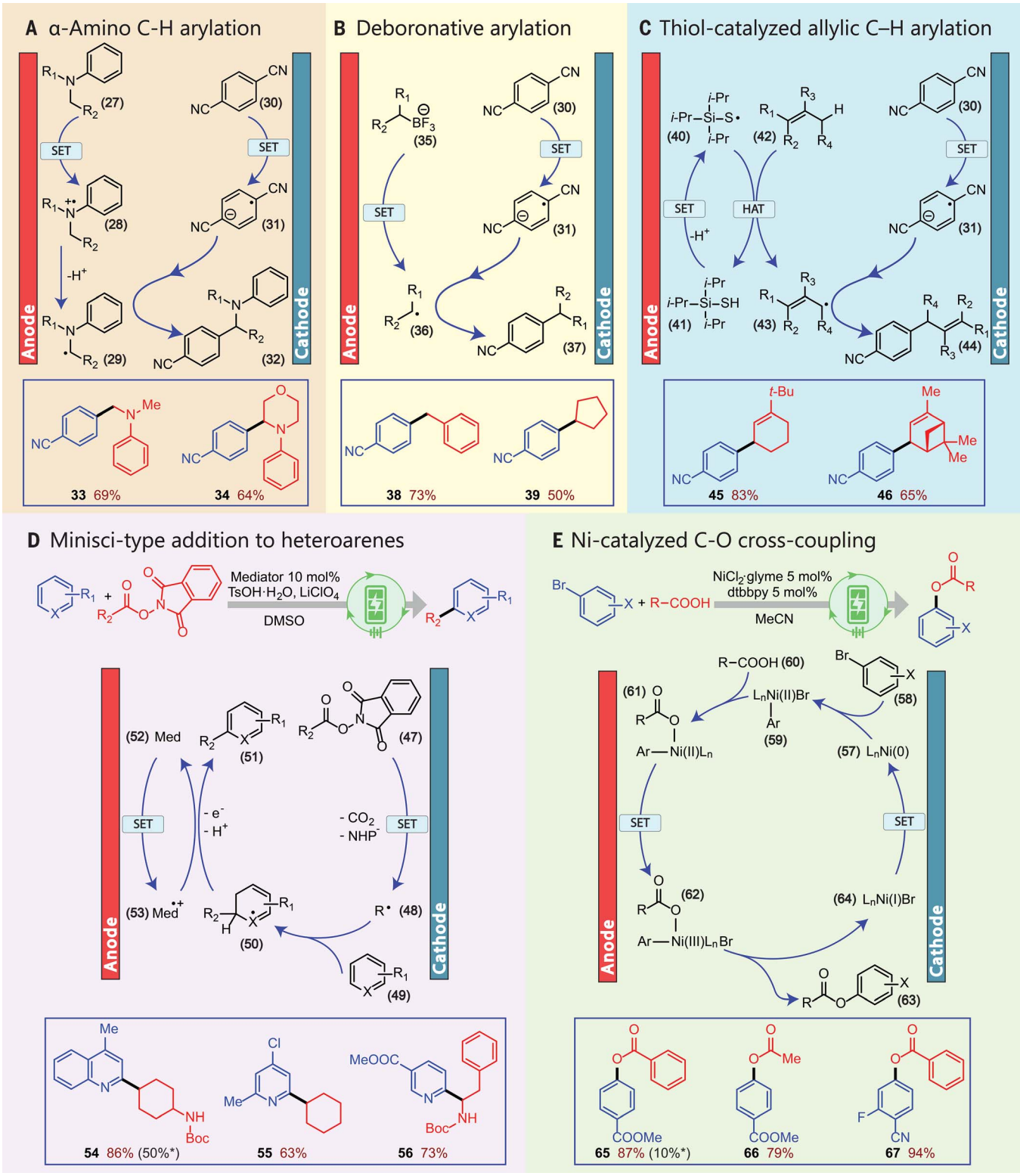


Fig. 3. General applicability of μRN-eChem platform for SET redox-neutral chemistry. (**A**) α-Amino C–H arylation. (**B**) Deboronative arylation. (**C**) Thiol-catalyzed allylic C–H arylation. HAT, hydrogen atom transfer; *i*-Pr, isopropyl. (**D**) Minisci-type radical addition to heteroarenes. Mediators (Med) used are ferrocene or 4-methoxytriphenylamine. Ts, tosyl; DMSO, dimethyl sulfoxide; NHP, phthalimide. (**E**) Ni-catalyzed C–O cross-coupling. dtbbpy, 4,4′-di-*tert*-butyl-2,2′-dipyridyl. Asterisks indicate high-performance liquid chromatography yield obtained from batch setup (fig. S5B) under identical electrochemistry conditions. See the supplementary materials for detailed reaction conditions.

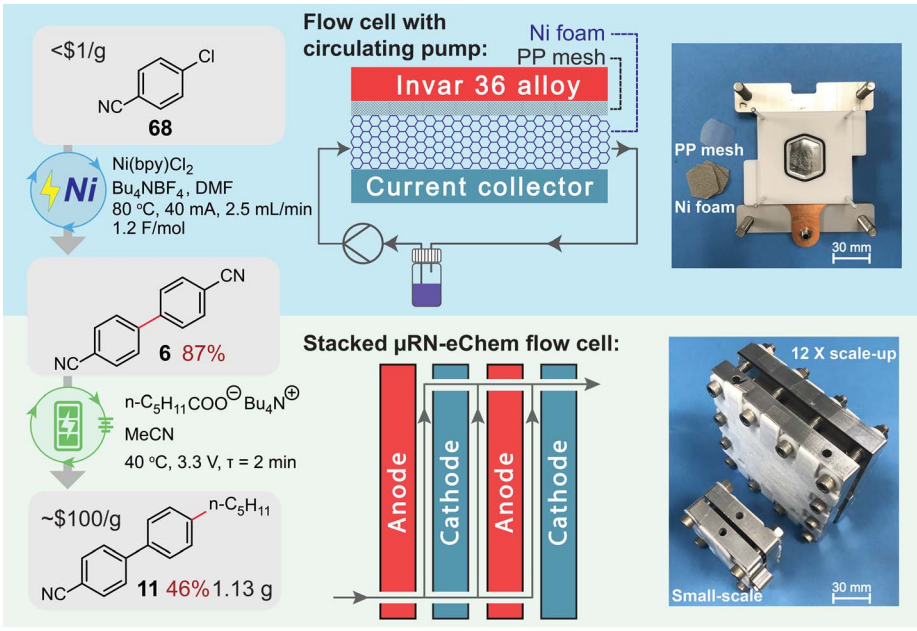


Fig. 4. A fully electrochemical two-step synthesis of liquid crystal material 5CB. See the supplementary materials for detailed reaction conditions. DMF, *N*,*N*′-dimethylformamide; PP, polypropylene.

In addition to the radical-radical cross-coupling, other types of SET redox-neutral transformations can also be implemented on our μRN-eChem platform. Recently, preactivated carboxylic acid *N*-hydroxyphthalimide esters have been used in photocatalyzed Minisci-type reactions to functionalize heterocycles, avoiding the high oxidation potential of directly activating carboxylic acids and thus improving functional group tolerance (*30*, *31*). In the presence of a redox mediator **52**, μRN-eChem is able to shuttle the electron between the anode and the transient radical **50** generated near the cathode, furnishing Minisci products in good yield (**54** to **56**, 63 to 86%) (Fig. 3D). Furthermore, μRN-eChem is compatible with SET redox-neutral transition-metal catalysis that has been extensively studied with photocatalysis (*32*). As an example, nickel-catalyzed C(sp^2)–O cross-coupling (*33*) was accomplished on the μRN-eChem platform to form O-aryl esters (**65** to **67**, 79 to 94% yield) (Fig. 3E). Mechanistically, the cathodic reduction of Ni(I) **64** to Ni(0) **57** facilitates the oxidative addition, and the anodic oxidation of Ni(II)(aryl) carboxylate **61** to Ni(III) **62** promotes the reductive elimination (*34*, *35*).

Compared with photochemistry, the catalyst-free, low-cost characteristics of μRN-eChem make it potentially more economical for synthesizing medium-value chemicals in addition to high-value fine chemicals (e.g., pharmaceuticals). We decided to demonstrate the practicality of μRN-eChem with the synthesis of 4-cyano-4′-pentylbiphenyl (5CB), a commonly used nematic liquid crystal material. Currently, 5CB is produced from biphenyl in a linear synthesis, which suffers from low yields and low efficiency (*36*). We proposed a fully electrochemical two-step synthesis to upgrade readily available 4-chlorobenzonitrile (4-ClBN) (<$1/g) to 5CB (~$100/g) with inexpensive reagents (Fig. 4). The first step, synthesis of BPDN, was a cathodic homocoupling of 4-ClBN catalyzed by $NiCl_2$ with 2,2′-bipyridine (bpy) as the ligand, and Invar *36* Fe/Ni alloy as the sacrificial anode. An electrochemical flow cell was engineered to handle gram-scale electrochemical synthesis (2.65 g synthesized in 24 hours) under air-free environment and elevated temperature, achieving an improved yield (87%) compared with a batch electrochemical setup (*37*). The second step used the μRN-eChem platform developed in this work to cross-couple BPDN and hexanoic acid under catalyst- and electrolyte-free conditions giving the targeted product (5CB). To demonstrate the scalability of μRN-eChem, we devised a three-layer stacked microfluidic flow cell (*38*, *39*), resulting in a 12-fold productivity increase relative to the small-scale flow cell. This scaled-up version continuously operated for 67 hours without intervention, yielding 1.13 g of 5CB.

The μRN-eChem strategy demonstrates that microfluidic technology and chemistry can work in tandem to controllably realize SET redox-neutral chemistry. Although the protocol uses an excess of one of the substrates, we anticipate that this conceptual approach will inspire the development of other redox-neutral electrosynthetic methods as a complement to existing redox-neutral photochemical synthesis technologies.

REFERENCES AND NOTES

1. J. M. R. Narayanam, C. R. J. Stephenson, *Chem. Soc. Rev.* **40**, 102–113 (2011).
2. M. H. Shaw, J. Twilton, D. W. C. MacMillan, *J. Org. Chem.* **81**, 6898–6926 (2016).
3. C. K. Prier, D. A. Rankic, D. W. C. MacMillan, *Chem. Rev.* **113**, 5322–5363 (2013).
4. N. A. Romero, D. A. Nicewicz, *Chem. Rev.* **116**, 10075–10166 (2016).
5. M.-C. Fu, R. Shang, B. Zhao, B. Wang, Y. Fu, *Science* **363**, 1429–1434 (2019).
6. C. J. O'Brien *et al.*, *J. Org. Chem.* **83**, 8926–8935 (2018).
7. J. J. Devery III *et al.*, *Chem. Sci.* **6**, 537–541 (2015).
8. N. V. Tsarevsky, K. Matyjaszewski, *Chem. Rev.* **107**, 2270–2299 (2007).
9. J. C. Theriot *et al.*, *Science* **352**, 1082–1086 (2016).
10. M. Yan, Y. Kawamata, P. S. Baran, *Chem. Rev.* **117**, 13230–13319 (2017).
11. Y. Kawamata *et al.*, *J. Am. Chem. Soc.* **141**, 6392–6402 (2019).
12. M. J. Llorente, B. H. Nguyen, C. P. Kubiak, K. D. Moeller, *J. Am. Chem. Soc.* **138**, 15110–15113 (2016).
13. M. D. Otero, B. Batanero, F. Barba, *Tetrahedron Lett.* **47**, 2171–2173 (2006).
14. C. Amatore, A. R. Brown, *J. Am. Chem. Soc.* **118**, 1482–1486 (1996).
15. Y. Ma *et al.*, *Angew. Chem. Int. Ed.* **58**, 16548–16552 (2019).
16. H. Kim *et al.*, *Science* **352**, 691–694 (2016).
17. D. Pletcher, R. A. Green, R. C. D. Brown, *Chem. Rev.* **118**, 4573–4591 (2018).
18. M. Atobe, H. Tateno, Y. Matsumura, *Chem. Rev.* **118**, 4541–4572 (2018).
19. J. Xie, H. Jin, A. S. K. Hashmi, *Chem. Soc. Rev.* **46**, 5193–5203 (2017).
20. H. Fischer, *Chem. Rev.* **101**, 3581–3610 (2001).
21. D. Leifert, A. Studer, *Angew. Chem. Int. Ed.* **59**, 74–108 (2020).
22. R. L. Hartman, J. P. McMullen, K. F. Jensen, *Angew. Chem. Int. Ed.* **50**, 7502–7519 (2011).
23. H. Kolbe, *J. Prakt. Chem.* **41**, 137–139 (1847).
24. A. McNally, C. K. Prier, D. W. C. MacMillan, *Science* **334**, 1114–1117 (2011).
25. Z. Zuo, D. W. C. MacMillan, *J. Am. Chem. Soc.* **136**, 5257–5260 (2014).
26. K. Nakajima, Y. Miyake, Y. Nishibayashi, *Acc. Chem. Res.* **49**, 1946–1956 (2016).
27. J. Suzuki, M. Tanigawa, S. Inagi, T. Fuchigami, *ChemElectroChem* **3**, 2078–2083 (2016).
28. A. J. J. Lennox, J. E. Nutting, S. S. Stahl, *Chem. Sci.* **9**, 356–361 (2017).
29. J. D. Cuthbertson, D. W. C. MacMillan, *Nature* **519**, 74–77 (2015).
30. R. S. J. Proctor, R. J. Phipps, *Angew. Chem. Int. Ed.* **58**, 13666–13699 (2019).
31. R. S. J. Proctor, H. J. Davis, R. J. Phipps, *Science* **360**, 419–422 (2018).
32. J. Twilton *et al.*, *Nat. Rev. Chem.* **1**, 0052 (2017).
33. E. R. Welin, C. Le, D. M. Arias-Rotondo, J. K. McCusker, D. W. C. MacMillan, *Science* **355**, 380–385 (2017).
34. L. Tian, N. A. Till, B. Kudisch, D. W. C. MacMillan, G. D. Scholes, *J. Am. Chem. Soc.* **142**, 4555–4559 (2020).
35. J. A. Terrett, J. D. Cuthbertson, V. W. Shurtleff, D. W. C. MacMillan, *Nature* **524**, 330–334 (2015).
36. P. J. Collings, M. Hird, in *Introduction to Liquid Crystals: Chemistry and Physics* (CRC Press, ed. 1, 1997), pp. 157–158.
37. R. Rahil, S. Sengmany, E. L. Gall, E. Léonel, *Synthesis* **02**, 146–154 (2018).
38. N. de Mas, A. Günther, T. Kraus, M. A. Schmidt, K. F. Jensen, *Ind. Eng. Chem. Res.* **44**, 8997–9013 (2005).
39. Y. Mo, J. Imbrogno, H. Zhang, K. F. Jensen, *Green Chem.* **20**, 3867–3874 (2018).

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Professor Jun-ichi Yoshida, who pioneered microfluidic electrochemistry and flash chemistry. We thank J. Raymond for the help with setup of the scale-up electrochemical synthesis experiment. We thank R. Y. Liu, A. W. Schuppe, C. P. Nguyen, and S. D. McCann for critical reading of the manuscript. **Funding:** This work was funded by the Novartis-MIT

Center for Continuous Manufacturing. Y.M. was supported by the Chyn Duog Shiah Memorial Fellowship from MIT. **Author contributions:** Y.M. proposed, designed, and developed microfluidic redox-neutral electrochemistry strategy. Y.M. and Z.L. designed and conducted the electrochemical single-electron redox-neutral reactions. Y.M., G.R., P.P., N.G., and A.I.A. designed and fabricated microfluidic electrochemical flow cells. S.L.B. supervised the chemistry development. K.F.J. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1352/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S25
Table S1
NMR Spectra
References (*40*–*55*)

27 November 2019; accepted 20 April 2020
10.1126/science.aba3823

GEOPHYSICS

3D fault architecture controls the dynamism of earthquake swarms

Zachary E. Ross[1]*, Elizabeth S. Cochran[2], Daniel T. Trugman[3,4], Jonathan D. Smith[1]

The vibrant evolutionary patterns made by earthquake swarms are incompatible with standard, effectively two-dimensional (2D) models for general fault architecture. We leverage advances in earthquake monitoring with a deep-learning algorithm to image a fault zone hosting a 4-year-long swarm in southern California. We infer that fluids are naturally injected into the fault zone from below and diffuse through strike-parallel channels while triggering earthquakes. A permeability barrier initially limits up-dip swarm migration but ultimately is circumvented. This enables fluid migration within a shallower section of the fault with fundamentally different mechanical properties. Our observations provide high-resolution constraints on the processes by which swarms initiate, grow, and arrest. These findings illustrate how swarm evolution is strongly controlled by 3D variations in fault architecture.

Earthquake swarms are enigmatic features of natural seismicity, with each sequence exhibiting a unique evolutionary pattern that is typically nonstationary in space and time. Swarms are frequently associated with regions of elevated heat flow (*1*, *2*) and are thought to be primarily driven by external processes such as fluids or aseismic slip (*3*–*5*) rather than earthquake triggering via stress transfer. The processes driving swarms are inferred from the observed diffusive or migratory evolutionary patterns and the tendency for the largest events to occur substantially after the onset of the sequence (*6*). However, much of the dynamism of swarms remains unexplained and likely reflects our limited knowledge of the structural and permeability architecture of the fault zones that produce them.

Fault zones are multiscale systems consisting of a localized fault core that accommodates the primary slip (*7*) and surrounding damage zones that contain highly fractured material. Damage zones commonly exhibit a fracture density that decays exponentially with distance from the fault core (*8*) and may contain additional smaller embedded structures (*9*). Fault zones have an appreciable influence on fluid flow in the crust, with higher-permeability damage zones acting as channels for fluids, whereas the fault core can act as either a conduit or a barrier to fluid flow (*10*, *11*). The permeability architecture of fault zones may be anisotropic (*12*) and transient (*13*, *14*), with geometric complexities sometimes creating secondary permeability structures in fault zones that further control fluid behavior (*15*).

[1]Seismological Laboratory, California Institute of Technology, Pasadena, CA, USA. [2]U.S. Geological Survey, Pasadena, CA, USA. [3]Department of Geological Sciences, The University of Texas at Austin, Austin, TX, USA. [4]Los Alamos National Laboratory, Los Alamos, NM, USA.
*Corresponding author. Email: zross@caltech.edu

Models for general fault architecture are predominantly compiled from geological and hydrological measurements obtained near the surface. Exhumed faults have advanced our understanding of the internal structure of fault zones at seismogenic depths (*7*, *16*) but provide limited constraints on their full three-dimensional (3D) variations. Occasionally, samples of fault material are obtained via drilling (*17*), which provide valuable but sparse constraints on deeper fault properties. Additionally, studies in mines provide glimpses of the complex faults and fracture surfaces that can occur at depth (*18*). These laterally heterogeneous structures are often generalized to three dimensions by assuming they vary smoothly with depth. However, recent seismological studies have increasingly observed that fault zones are highly 3D systems with complex depth-dependent variations (*19*, *20*), numerous strands over many scales (*21*, *22*), and intricate spatially variable damage structures (*23*). These observations provide insights into the connections between fault structure and earthquake physics, yet gaps remain in understanding deep fault structure and permeability architecture, which is essential for understanding the complex evolution of earthquake swarms.

From early 2016 to late 2019, a long-term, vigorous swarm occurred near Cahuilla, California, roughly halfway between the San Jacinto and Elsinore fault zones (*24*) (Fig. 1). The swarm likely occurred in the Cahuilla Valley pluton, a hornblende-biotite tonalite that is the predominant intrusive rock in the area (*25*). To better understand the sequence and properties of the fault zone, we used earthquake detection algorithms with deep neural networks to produce a highly detailed seismicity catalog of more than 22,000 events with moment magnitudes ranging from 0.7 to 4.4 (*26*). We discovered a fault zone with rich 3D structure that controls the dynamic evolution of the swarm.

A high-resolution perspective of the 3D geometry of the fault zone is observed in the relocated seismicity. The swarm occurred in a single nonplanar right-lateral strike-slip fault surface that dips 70° to 80° toward the northeast (Fig. 1). In the fault-normal direction, the seismicity is just tens of meters thick, which is comparable to the relative horizontal location errors (38 m); the seismogenic portion of the fault zone may, therefore, be even narrower. The fault geometry is nonplanar both along strike and along dip, most notably at the northwestern edge, where the fault surface undergoes multiple reversals in the dip direction as a function of depth.

Within the fault zone, there is a stack of elongated strike-parallel channels of relatively high seismicity density with vertical separation of ~200 to 400 m. This distance is larger than the relative vertical error (87 m). These channels have a weakly undulating geometry and persist over most of the fault zone. They are also prominently observed in the aggregate distribution of focal depths (Fig. 1), where they are represented by five peaks.

The 3D fault geometry is critical for understanding the spatiotemporal evolution of the swarm (Fig. 2). We can trace the origin of the sequence to a 100-m–wide region near the base of the fault zone (8-km depth). The swarm expanded slowly in the fault zone over several years, ultimately extending across an area that measured 4 km by 4 km. We observed a prominent front to the migration, with sustained earthquake activity behind it (Fig. 2). However, virtually no earthquake triggering occurs ahead of the front, which is consistent with expectations for a sustained point injection source (*27*). By interpreting the migration of seismicity as due to fluid flow, the slow diffusion front suggests that the swarm is driven predominantly by fluid migration (*28*) rather than static or dynamic triggering, as is more typical of earthquake sequences. The migration speeds are extremely slow at 1 to 5 m/day, which is much slower than the propagation speed of slow slip events (tens of kilometers per day) (*3*) and also slower than many natural injection-driven swarms (*29*). We estimate the permeability of this fault zone to be ~10^{-17} to 10^{-18} m^2 (*26*). The observations bear similarities to earthquake sequences induced by deep fluid disposal (*27*), although here fluid injection is occurring directly into the fault zone from a deep, natural source.

The spatial expansion of the swarm is not uniform, indicating that the permeability architecture of the fault zone is variable (Fig. 2, A and B). There may be a permeability barrier 0.5 km below the injection source at around 8-km depth, as only a minimal amount of deeper seismicity occurred during the nearly 4 years of activity. This may alternatively indicate that the fault zone itself does not extend

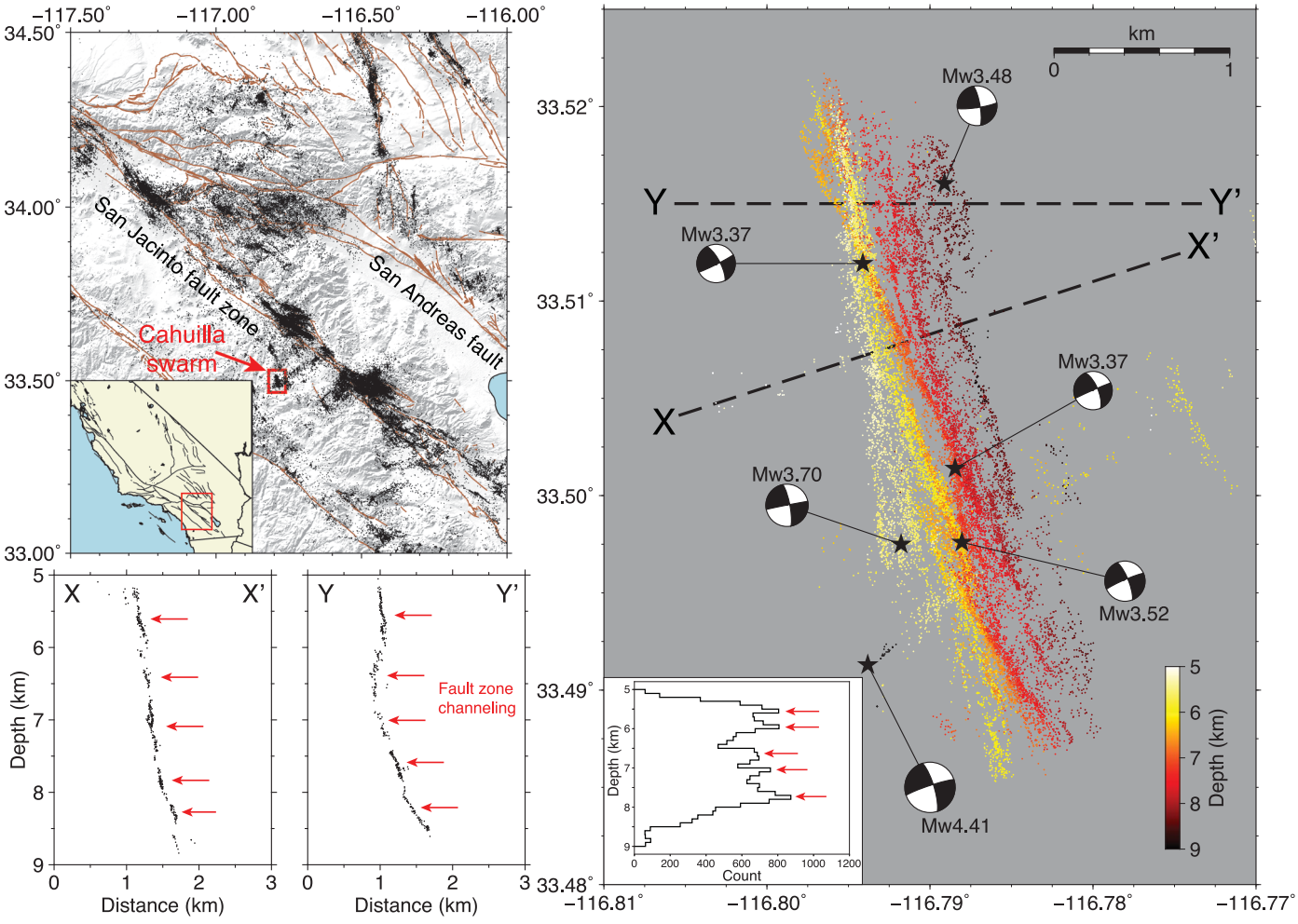


Fig. 1. Relocated seismicity during the Cahuilla swarm. Fault zone is less than 50 m wide and exhibits prominent channeling along strike. Cross sections X-X' and Y-Y' show that there is considerable curvature to the fault geometry over short distances.

to greater depths or that there was a change in rheology at this point. Laterally, the swarm was constrained on both sides after expanding to an along-strike length of 4 km, which was reached ~3 years into the sequence. This lateral expansion was asymmetric, migrating ~1.8 times faster to the northwest than to the southeast. Further expansion may be limited by permeability barriers, or the fault zone itself may not extend beyond the final spatial extent of the sequence. Fluid pressures may also no longer be high enough to produce additional events. In the fault-normal direction, the swarm is limited to a few tens of meters in width, with no evidence of expansion (Fig. 1). Our observations are consistent with models of fault architecture that have a damage zone embedded in relatively impermeable host rock (*11*).

From 2016 to mid-2017, the migration velocity of the swarm was roughly isotropic (Fig. 2C). Then, the along-strike component of the migration accelerated relative to the along-dip component, which coincided with the fluids entering the fault channels.

In August 2018, nearly 3 years after the swarm's onset, the up-dip progression stalled at a depth of 6 km across most of the fault zone (Fig. 2B). Careful examination of the seismicity indicates that up-dip migration still continues, but only through a 1-km-long strike-parallel aperture near the southeast end of the fault zone (Fig. 2B). After the localized events migrated upwards for 500 m, seismic activity escalated rapidly and eventually culminated in the largest earthquake of the sequence, a moment magnitude (M_w) 4.4 event. The second-largest event to precede the M_w 4.4 mainshock was a M_w 2.8 earthquake that occurred 4 days earlier. After the mainshock, the swarm migrated almost unilaterally to the northwest, filling a gap in seismicity at depths less than 6 km. From these observations, we infer a 2-km-long subhorizontal permeability barrier at 6-km depth, with limited to no fluid flow across it. This barrier correlates with a dip reversal of the fault zone (Fig. 1, Y-Y'). Thus, permeability structures in the fault zone exert a major control on the local diffusion properties.

Approximately 850 days into the sequence, the fluid injection may stop, as evidenced by the lack of seismicity within 1 km of the injection source, and swarm migration continues at further distances (Fig. 2C).

The M_w 4.4 earthquake denotes a key transition point in the swarm. The sequence initiated with steady but low seismicity rates from 2016 to 2017, followed by somewhat higher rates until the occurrence of the M_w 4.4 earthquake (Fig. 2C, periods I and II). Immediately after the M_w 4.4 earthquake, the seismicity rate increased substantially (Fig. 2C, period III) but then decayed, with the swarm eventually dying out about a year later. The seismicity rate throughout period III was much more typical for earthquake swarms than the steady rate observed in periods I and II, as there is stronger time dependence and temporal clustering (*6*).

The distinctly different character of periods I and II compared with period III leads to the question of what physical processes control the event rate behavior. The simplest explanation is that during periods I and II, the swarm

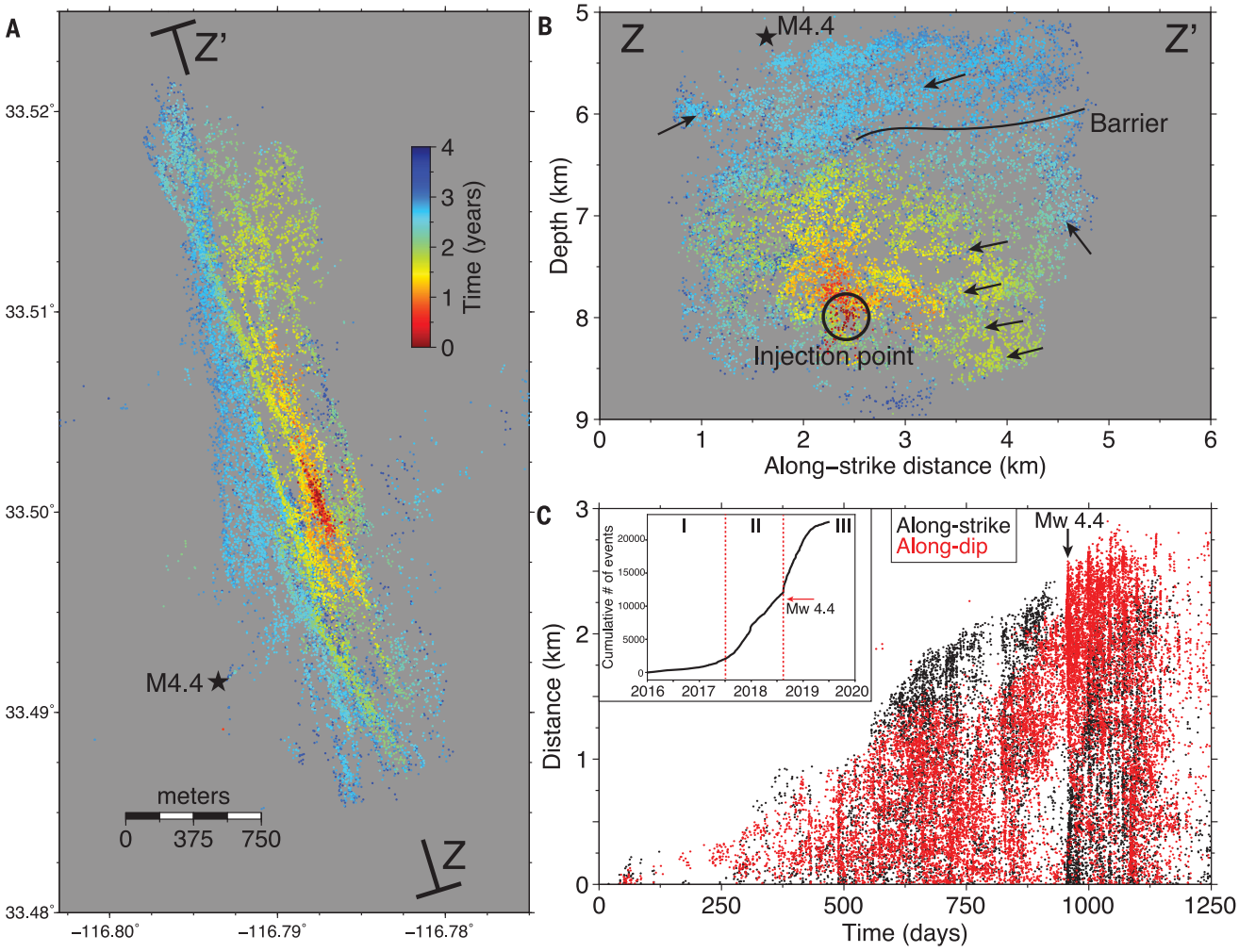


Fig. 2. Spatiotemporal evolution of the sequence. Fluids are naturally injected into the base of the fault zone and diffuse within it through channels (black arrows). (**A**) Map view of seismicity evolution, with events color-coded by occurrence time. (**B**) Fault-parallel cross section along Z-Z', including the inferred permeability barrier and injection point. (**C**) Absolute distance of seismicity relative to injection point in (B) projected into along-strike and along-dip components.

was driven predominantly by fluid pressure diffusion through the fault zone, whereas during period III, the swarm was driven by a mixture of static stress changes and fluid pressure changes. From Fig. 2C, periods I and II exhibit no triggering ahead of the front and little clustering, whereas period III has pervasive clustering. The M_w 4.4 earthquake produced appreciable stress changes on the surrounding fault zone that may have altered the permeability structure of the damage zone, possibly enhancing fluid flow through the system.

We examined earthquake stress drops ($\Delta\sigma$) in the context of the structural architecture of the fault zone and sequence evolution. Stress drop is a fundamental earthquake source parameter and a key component of the seismic energy budget (*30*). We derived $\Delta\sigma$ for earthquakes with local magnitude (M_L) > 0.5 from a spectral decomposition approach (*26*). Figure 3A shows the median $\Delta\sigma$ in a sliding window over the duration of the sequence. Throughout most of period II, the median stress drop fluctuates around 8 MPa (Fig. 3A) before dropping markedly to a range of 2 to 3 MPa. The median $\Delta\sigma$ remains low for ~5 months before a stepwise recovery to a value that approaches the pre-mainshock level.

Before the occurrence of the M_w 4.4 earthquake (period II), the swarm located almost entirely below 6-km depth, and the $\Delta\sigma$ values were generally in the range 4 to 15 MPa (median ~8 MPa) (Fig. 3B). As the fluids circumvented the barrier at 6-km depth, coincident with the M_w 4.4 earthquake, the median values of $\Delta\sigma$ showed a prominent step decrease to around 2 MPa. Fluids then moved horizontally through a channel toward the northwest (updip of the 6-km barrier), and in this channel, $\Delta\sigma$ values were persistently lower (1 to 5 MPa). The $\Delta\sigma$ of earthquakes down-dip of the 6-km depth barrier during this time were in the same range as they were before the M_w 4.4 earthquake. The pronounced change in median $\Delta\sigma$ at 6-km depth, along with the inferred permeability barrier at the same place, suggests a sharp interface between materials with fundamentally different structural properties.

We used deep-learning algorithms to produce detailed images of fault zone structure at seismogenic depths, which reveal a system with permeability that is highly variable in three dimensions. In particular, we resolve strike-parallel channeling in the fault zone that controls the migration of events. These observations may relate to channeling seen in numerical and laboratory experiments of flow in fault zones and fractures (*31*, *32*), or to anisotropic fault roughness (*33*). We also resolve a subhorizontal permeability barrier that, when circumvented, allows for the largest earthquake in the sequence. Together, these factors illustrate how the 3D permeability structure of the fault zone controlled the evolution of the Cahuilla swarm (Fig. 4).

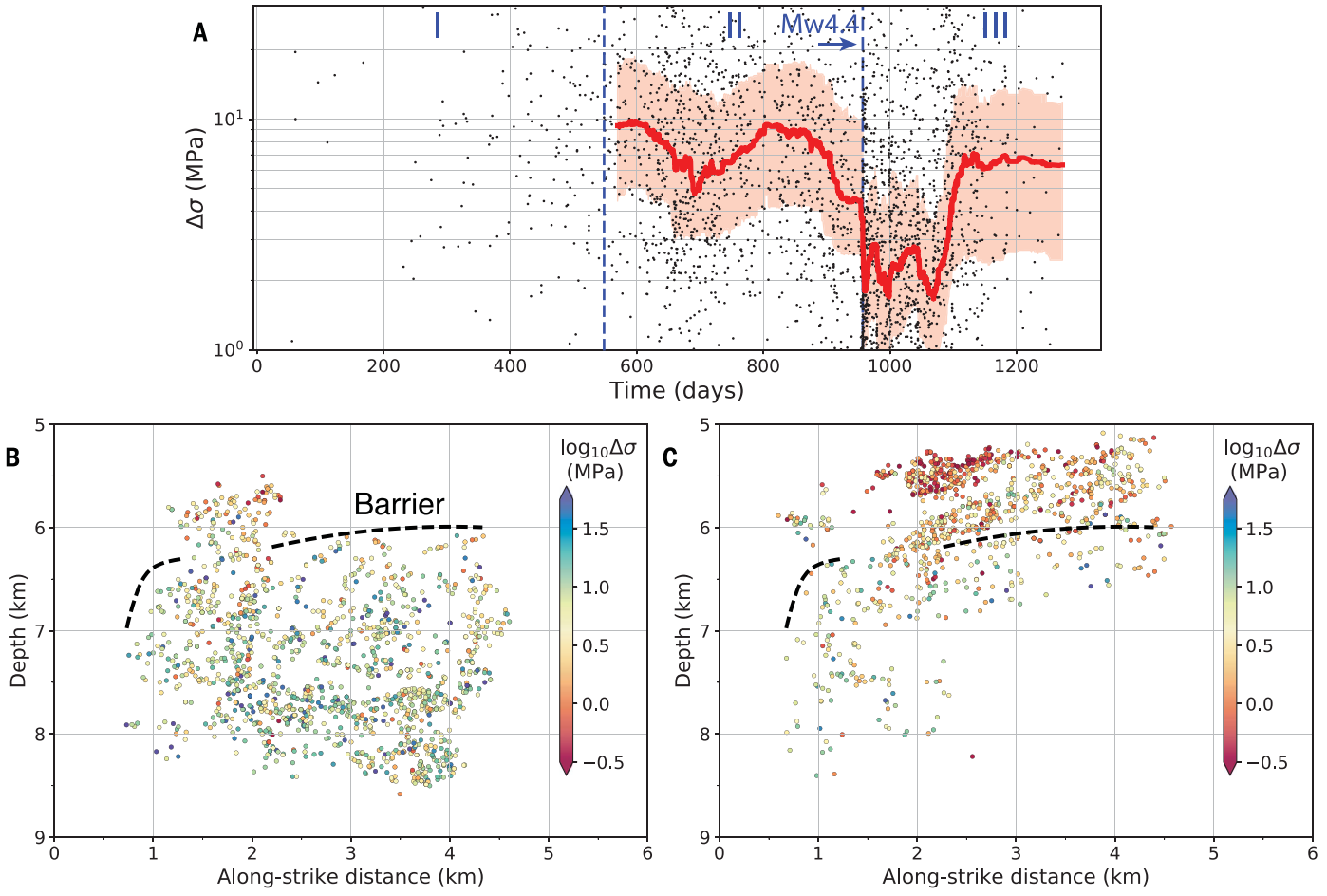


Fig. 3. Evolution stress drop (Δσ). (**A**) Time evolution. The solid red line marks the median Δσ in a causal, 200-event moving window. Shading denotes the interquartile range. (**B**) Along-strike cross-sectional view of events occurring before the M_w 4.4 earthquake. Bluer colors indicate events that are enriched in high-frequency energy. Δσ decreases as the swarm circumvents the barrier at 6-km depth. (**C**) Along-strike cross-sectional view of events occurring in the first 50 days after the M_w 4.4 earthquake. For events shallower than 6 km, Δσ is an order of magnitude lower.

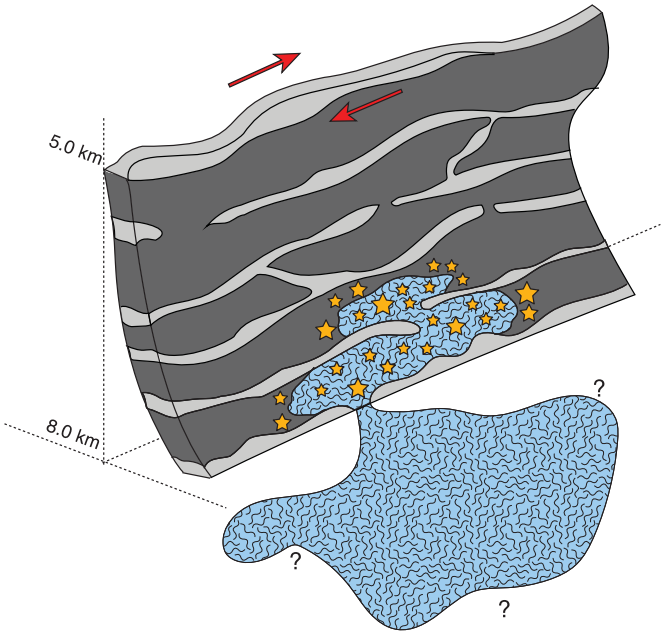


Fig. 4. Cartoon summarizing observations. A fault zone is embedded in host rock and exhibits geometric warping along strike and along dip. Within the fault zone, there are undulating subhorizontal channels. A small portion of the base of the fault zone connects to a deeper reservoir that is initially sealed. Once the seal breaks, fluid is injected into the fault zone, which diffuses primarily through channels, driving the seismicity. During the nearly 4 years of activity, fluids interact with numerous permeability barriers.

Standard models for fault zone hydrogeology mainly distinguish between fault-parallel and fault-normal flow (*11*) and how these systems act as conduits and/or barriers to flow depending on the geometrical aspects of the damage zone. Although our results provide evidence in support of these models, particularly at seismogenic depths, for which previous observations are lacking, they also highlight limitations in generalizing fault architecture to three dimensions from near-surface observations alone.

The prominent diffusive character of the swarm relative to a highly localized injection point suggests that the fault zone connects to a reservoir at depth but was sealed initially. This scenario is described by the fault valve model (*34*), in which a permeability seal at the roots of a fault zone is suddenly broken, allowing fluid to be transported up-dip through the fault zone via an overpressured hydraulic gradient. Fault valving has been invoked to explain the activation of faults under various circumstances (*19*, *29*, *35*), but most of the evidence

for this phenomenon has been circumstantial. Our observations provide clear evidence of fault valving as a mechanism for seismogenesis in swarms. This is noteworthy because understanding how and why earthquake sequences initiate is among the greatest outstanding challenges in earthquake science. Because two nearby swarms in the 1980s also exhibited migration patterns over several years (*24*), it is likely that this reservoir has a large spatial extent and connects to many faults in the area, making this a key mechanism for transporting fluids through the crust.

Permeability is not a static, scalar property but instead a tensor that may vary in space and time. The Cahuilla swarm neatly illustrates this point, as the spatiotemporal evolution of seismicity is quite different from what would be observed if permeability were constant and isotropic. Permeability is expected to evolve with time in response to slip and fault healing processes, a fact that has important implications for earthquake cycle dynamics (*29*, *36*). Improved observational constraints of these processes as they occur in real fault systems will, therefore, help inform our understanding of earthquake physics. Measurements of earthquake stress drop commonly exhibit variability of 3 to 4 orders of magnitude (*37*), and understanding how much of this is natural (rather than due to error) is important because the variability can be related back to the intrinsic structural properties of fault zones or extrinsic factors such as stress. We observe a sudden change of nearly an order of magnitude in the median stress drops of events in the upper 1 km of the sequence. By having detailed knowledge of the internal structure of the fault zone that produced the Cahuilla swarm, we conclude that the fault is divided vertically into two sections with different mechanical properties that are separated by a permeability barrier.

Earthquake science has always been challenged by the staggering evolutionary complexity observed within and between different sequences. This complexity has made it difficult to distill common characteristics and formulate generalized conclusions that can provide constraints on the physics of earthquakes. With improving resolution, however, we are able to observe that sequence complexity often arises from the 3D characteristics of fault zones, which are themselves dynamic.

REFERENCES AND NOTES

1. J. E. Vidale, P. M. Shearer, *J. Geophys. Res.* **111**, 12 (2006).
2. I. Zaliapin, Y. Ben-Zion, *J. Geophys. Res.* **118**, 2865–2877 (2013).
3. E. Roland, J. J. McGuire, *Geophys. J. Int.* **178**, 1677–1690 (2009).
4. S. Hainzl, T. Fischer, H. Čermáková, M. Bachura, J. Vlček, *J. Geophys. Res.* **121**, 2575–2590 (2016).
5. S. A. Miller *et al*., *Nature* **427**, 724–727 (2004).
6. X. Chen, P. M. Shearer, *J. Geophys. Res.* **116**, B09309 (2011).
7. F. M. Chester, J. S. Chester, *Tectonophysics* **295**, 199–221 (1998).
8. T. M. Mitchell, D. R. Faulkner, *J. Struct. Geol.* **31**, 802–816 (2009).
9. Y. Ben-Zion, C. G. Sammis, *Pure Appl. Geophys.* **160**, 677–715 (2003).
10. J. S. Caine, J. P. Evans, C. B. Forster, *Geology* **24**, 1025–1028 (1996).
11. V. F. Bense, T. Gleeson, S. E. Loveless, O. Bour, J. Scibek, *Earth Sci. Rev.* **127**, 171–192 (2013).
12. V. F. Bense, M. A. Person, *Water Resour. Res.* **42**, W05421 (2006).
13. M. Manga *et al*., *Rev. Geophys.* **50**, RG2004 (2012).
14. L. Xue *et al*., *Science* **340**, 1555–1559 (2013).
15. R. H. Sibson, *J. Struct. Geol.* **18**, 1031–1042 (1996).
16. D. R. Faulkner, A. C. Lewis, E. H. Rutter, *Tectonophysics* **367**, 235–251 (2003).
17. F. M. Chester *et al*., *Science* **342**, 1208–1211 (2013).
18. V. Heesakkers, S. Murphy, Z. Reches, *Pure Appl. Geophys.* **168**, 2395–2425 (2011).
19. D. R. Shelly, W. L. Ellsworth, D. P. Hill, *J. Geophys. Res.* **121**, 1776–1795 (2016).
20. A. A. Allam, Y. Ben-Zion, *Geophys. J. Int.* **190**, 1181–1196 (2012).
21. S. J. Wei *et al*., *Nat. Geosci.* **4**, 615–618 (2011).
22. Z. E. Ross *et al*., *Science* **366**, 346–351 (2019).
23. Z. E. Ross, E. Hauksson, Y. Ben-Zion, *Sci. Adv.* **3**, e1601946 (2017).
24. E. Hauksson, Z. E. Ross, E. Cochran, *J. Geophys. Res.* **124**, 3933–3949 (2019).
25. R. V. Sharp, *Bull. Geol. Soc. Am.* **78**, 705–730 (1967).
26. Materials and methods are described in the supplementary materials.
27. S. A. Shapiro, P. Audigane, J.-J. Royer, *Geophys. J. Int.* **137**, 207–213 (1999).
28. S. Hainzl, Y. Ogata, *J. Geophys. Res.* **110**, B05S07 (2005).
29. S. F. Cox, *Econ. Geol.* **111**, 559–587 (2016).
30. H. Kanamori, L. Rivera, in *Earthquakes: Radiated Energy and the Physics of Faulting*, R. Abercrombie, A. McGarr, G. D. Toro, H. Kanamori, Eds. (American Geophysical Union, 2006), pp. 3–13.
31. N. E. Odling, S. D. Harris, R. J. Knipe, *J. Struct. Geol.* **26**, 1727–1747 (2004).
32. S. Brown, A. Caprihan, R. Hardy, *J. Geophys. Res.* **103**, 5125–5132 (1998).
33. W. L. Power, T. E. Tullis, S. R. Brown, G. N. Boitnott, C. H. Scholz, *Geophys. Res. Lett.* **14**, 29–32 (1987).
34. R. H. Sibson, in *Earthquake Prediction: An International Review*, vol. 4, D. W. Simpson, P. G. Richards, Eds. (American Geophysical Union, 1981), pp. 593–603.
35. R. H. Sibson, *Earth Planet. Sci. Lett.* **257**, 188–199 (2007).
36. J. M. Gosselin *et al*., *Sci. Adv.* **6**, eaay5174 (2020).
37. P. M. Shearer, G. A. Prieto, E. Hauksson, *J. Geophys. Res.* **111**, B06303 (2006).
38. California Institute of Technology Dataset, Southern California Earthquake Data Center (2013); doi:10.7909/C3WD3xH1.

ACKNOWLEDGMENTS

Funding: Z.E.R. was supported by the Southern California Earthquake Center (award 20018). D.T.T. acknowledges institutional support from the Laboratory Directed Research and Development (LDRD) program of Los Alamos National Laboratory (project number 20180700PRD1). **Author contributions:** Z.E.R. constructed the seismicity catalog; Z.E.R. and E.S.C. performed the spatiotemporal seismicity analysis; D.T.T. performed the magnitude and stress drop calculations; all authors contributed to general analysis and the writing of the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data used in this study are publicly available from the Southern California Seismic Network (*38*). The produced seismicity catalog will be publicly hosted at the Southern California Earthquake Data Center (*38*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1357/suppl/DC1
Materials and Methods
Figs. S1 to S6
References (*39*–*54*)

28 January 2020; accepted 28 April 2020
10.1126/science.abb0779

CORONAVIRUS

Rapid implementation of mobile technology for real-time epidemiology of COVID-19

David A. Drew[1]*, Long H. Nguyen[1]*, Claire J. Steves[2,3], Cristina Menni[2], Maxim Freydin[2], Thomas Varsavsky[4], Carole H. Sudre[4], M. Jorge Cardoso[4], Sebastien Ourselin[4], Jonathan Wolf[5], Tim D. Spector[2,5]†, Andrew T. Chan[1,6]†‡, COPE Consortium§

The rapid pace of the coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) presents challenges to the robust collection of population-scale data to address this global health crisis. We established the COronavirus Pandemic Epidemiology (COPE) Consortium to unite scientists with expertise in big data research and epidemiology to develop the COVID Symptom Study, previously known as the COVID Symptom Tracker, mobile application. This application—which offers data on risk factors, predictive symptoms, clinical outcomes, and geographical hotspots—was launched in the United Kingdom on 24 March 2020 and the United States on 29 March 2020 and has garnered more than 2.8 million users as of 2 May 2020. Our initiative offers a proof of concept for the repurposing of existing approaches to enable rapidly scalable epidemiologic data collection and analysis, which is critical for a data-driven response to this public health challenge.

The exponentially increasing number of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections has led to "an urgent need to expand public health activities in order to elucidate the epidemiology of the novel virus and characterize its potential impact" (*1*). Understanding risk factors for infection and predictors of subsequent outcomes is crucial to gain control of the coronavirus disease 2019 (COVID-19) pandemic (*2*). However, the speed at which the pandemic is unfolding poses an unprecedented challenge for the collection of exposure data to characterize the full breadth of disease severity, hampering efforts for timely dissemination of accurate information to affect public health planning and clinical management. Thus, there is an urgent need for an adaptable real-time data-capture platform to rapidly and prospectively collect actionable high-quality data that encompass the spectrum of subclinical and acute presentations and identify disparities in diagnosis, treatment, and clinical outcomes. Addressing this priority will allow for more accurate estimates of disease incidence, inform risk mitigation strategies, facilitate allocation of scarce testing resources, and encourage appropriate quarantine and treatment of those afflicted.

An evolving body of literature suggests that COVID-19 incidence and outcomes vary according to age, sex, race, ethnicity, and underlying health status, with inconsistent evidence suggesting that commonly used medications—such as angiotensin-converting enzyme inhibitors, thiazolidinediones, and ibuprofen—may alter the natural disease course (*3*–*9*). Further, symptoms of COVID-19 vary widely, with fever and dry cough reportedly the most prevalent, though numerous investigations have demonstrated that asymptomatic carriage is a significant determinant of community spread (*5*–*7*, *10*–*13*). In addition, the full spectrum of clinical presentation, which is still being characterized, may differ markedly across patient subgroups. Recent advisories from the American Gastroenterological Association, the American Academy of Otolaryngology–Head and Neck Surgery, and the British Geriatrics Society have emphasized the potential association between COVID-19 infection and previously underappreciated gastrointestinal symptoms (e.g., nausea, anorexia, and diarrhea), loss of taste and/or smell, and common geriatric syndromes (e.g., falls and delirium). The pandemic has considerably outpaced our collective efforts to fully characterize who is most at risk or may suffer the most serious sequelae of infection.

Mobile phone applications and web-based tools facilitate self-guided collection of population-level data at scale (*14*), the results of which can be rapidly redeployed to inform participants of urgent health information (*14*, *15*). Both technologies are particularly advantageous when many individuals are advised to maintain physical distance from others (*16*). Such digital tools have already been applied in more controlled research settings, and these studies benefit from greater lead time for field testing, question curation, and recruitment. Although many digital collection tools for COVID-19 are being developed and launched in the Unites States and abroad (see http://mhealth-hub.org/mhealth-solutions-against-covid-19 for a continuously updated resource list from the European Union and the World Health Organization), including some in partnership with government health agencies such as the Centers for Disease Control and Prevention, most applications have largely been configured to offer a single assessment of symptoms to tailor semipersonalized recommendations for further evaluation. Infectious disease surveillance web-based tools (e.g., http://flunearyou.org) have been rapidly adapted for COVID-19–specific collection (e.g., http://covidnearyou.org). Alternatively, web portals have been developed for researchers to report patient-level information on behalf of participants already enrolled in clinical registries (e.g., ccc19.org). Integration with approaches that use remote data capture (e.g., wearable technology or symptom checkers such as real-time reporting thermometers) is also being considered. Although these approaches offer critical public health insights, they are often not tailored for the type of scalable longitudinal data capture that epidemiologists need to perform comprehensive, well-powered investigations.

To meet this challenge, we established a multinational collaboration, the COronavirus Pandemic Epidemiology (COPE) Consortium, composed of leading investigators from several large clinical and epidemiological cohort studies. COPE brings together a multidisciplinary team of scientists with expertise in big data research and translational epidemiology to investigate the COVID-19 pandemic in a large and diverse patient population. Several large cohorts have already agreed to join these efforts, including the Nurses' Health Study series, the Growing Up Today Study (GUTS), the Health Professionals Follow-Up Study (HPFS), TwinsUK, American Cancer Society Cancer Prevention Study 3 (CPS-3), the Multiethnic Cohort Study, the California Teachers Study (CTS), the Black Women's Health Study (BWHS), the Sister Study, Aspirin in Reducing Events in the Elderly (ASPREE), the Stanford Nutrition Studies, the Gulf Long-term Follow-up (GuLF) Study, the Agricultural Health Study, the National Institute of Environmental Health Sciences Environmental Polymorphisms Registry, the Predicting Progression of Developing Myeloma in a High-Risk Screened Population (PROMISE) study, and the Precursor Crowdsourcing (PCROWD) study. To aid in our data harmonization efforts in the United States, we developed the COVID Symptom Study (previously known as the COVID

[1]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, 55 Fruit St., Boston, MA 02114, USA. [2]Department of Twin Research and Genetic Epidemiology, King's College London, Westminster Bridge Road, London SE1 7EH, UK. [3]Department of Ageing and Health, Guys and St. Thomas' NHS Foundation Trust, Lambeth Palace Road, London SE1 7EH, UK. [4]School of Biomedical Engineering & Imaging Sciences, King's College London, 1 Lambeth Palace Road, London SE1 7EU, UK. [5]Zoe Global Limited, 164 Westminster Bridge Road, London SE1 7RW, UK. [6]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, 665 Huntington Ave., Boston, MA 02114, USA.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: achan@mgh.harvard.edu
§COPE Consortium members and affiliations are listed in the supplementary materials.

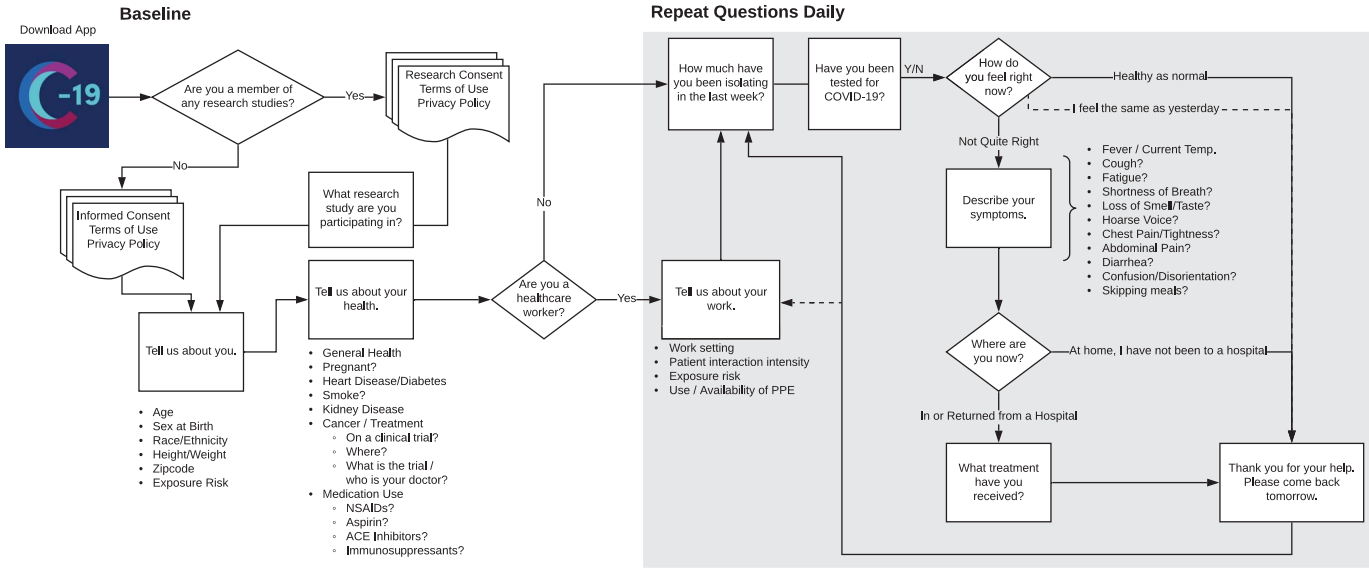


Fig. 1. Schematic of the participant workflow. After downloading the COVID Symptom Study app and providing consent, users are prompted to enter baseline demographic and clinical information and are serially queried about new or ongoing symptoms, testing results, and extent of isolation. Health care workers offer additional information about the intensity of their patient interactions, potential exposure to infected patients, and use of PPE. With informed consent, users also participating in a variety of ongoing cohorts or clinical trials (Nurses' Health Study, TwinsUK, and others) have the option of linking COVID Symptom Study information to their extant research data.

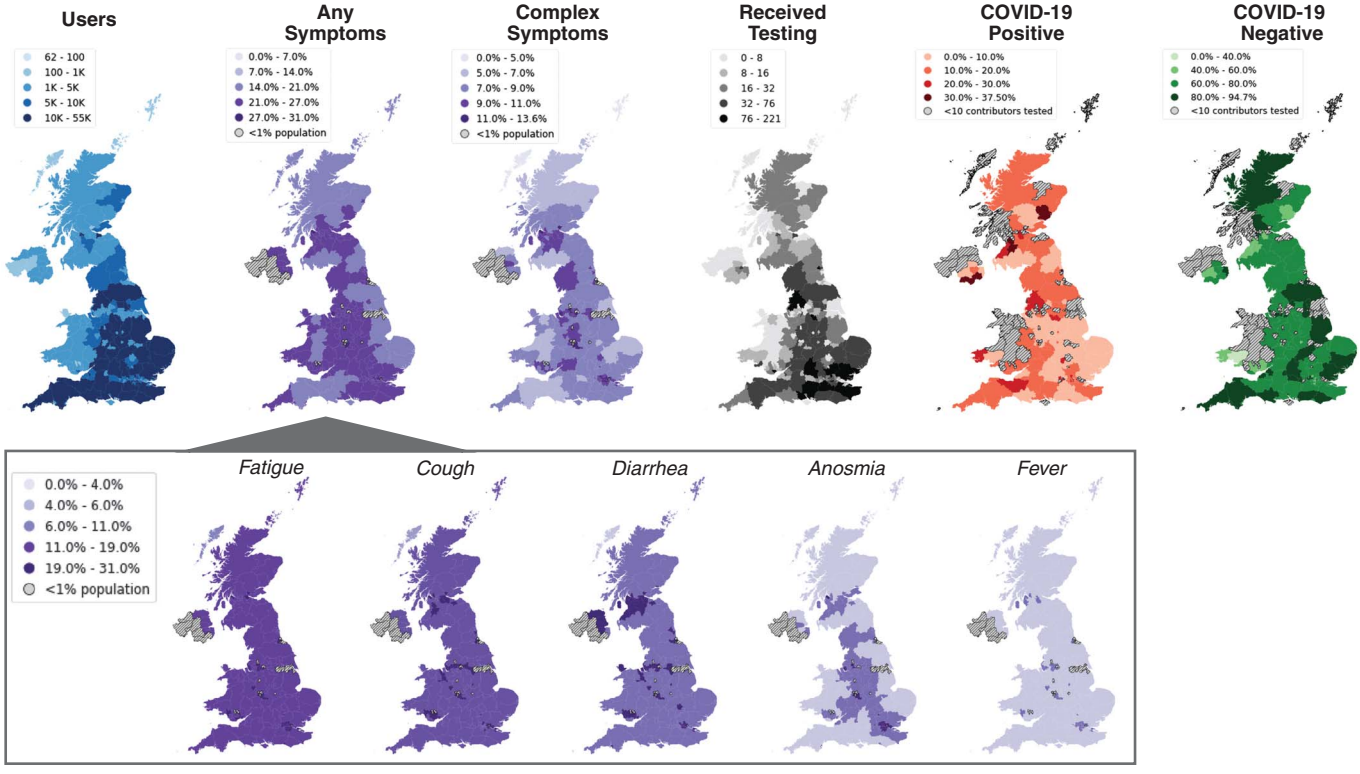


Fig. 2. COVID Symptom Study use, reported symptoms, and testing results according to geographic location in the United Kingdom. Between 24 and 29 March 2020, more than 1.6 million unique individuals downloaded the application and shared clinical and demographic information, as well as daily symptoms and high-intensity occupational exposures (blue map). Population density of those presenting with any symptoms (left purple map) varied according to region, with widespread reports of fatigue, cough, and diarrhea, followed by anosmia and, relatively infrequently, fever (inset). Examination of individuals who reported complex symptoms (right purple map), defined as having cough or fever and at least one other symptom (diarrhea, anosmia, or fever), reveals areas that potentially need more testing. For the subset of the population that received a COVID-19 test (black map), areas with larger proportions of positive tests (orange map) appear to coincide with areas in which high proportions of the population reported complex symptoms. By contrast, some areas with low prevalence of complex symptoms have received higher rates of testing and, consequently, more negative tests (green map). This example of real-time visualization of data captured by the COVID Symptom Study may help public health and government officials reallocate resources, identify areas with unmet testing needs, and detect emerging hotspots.

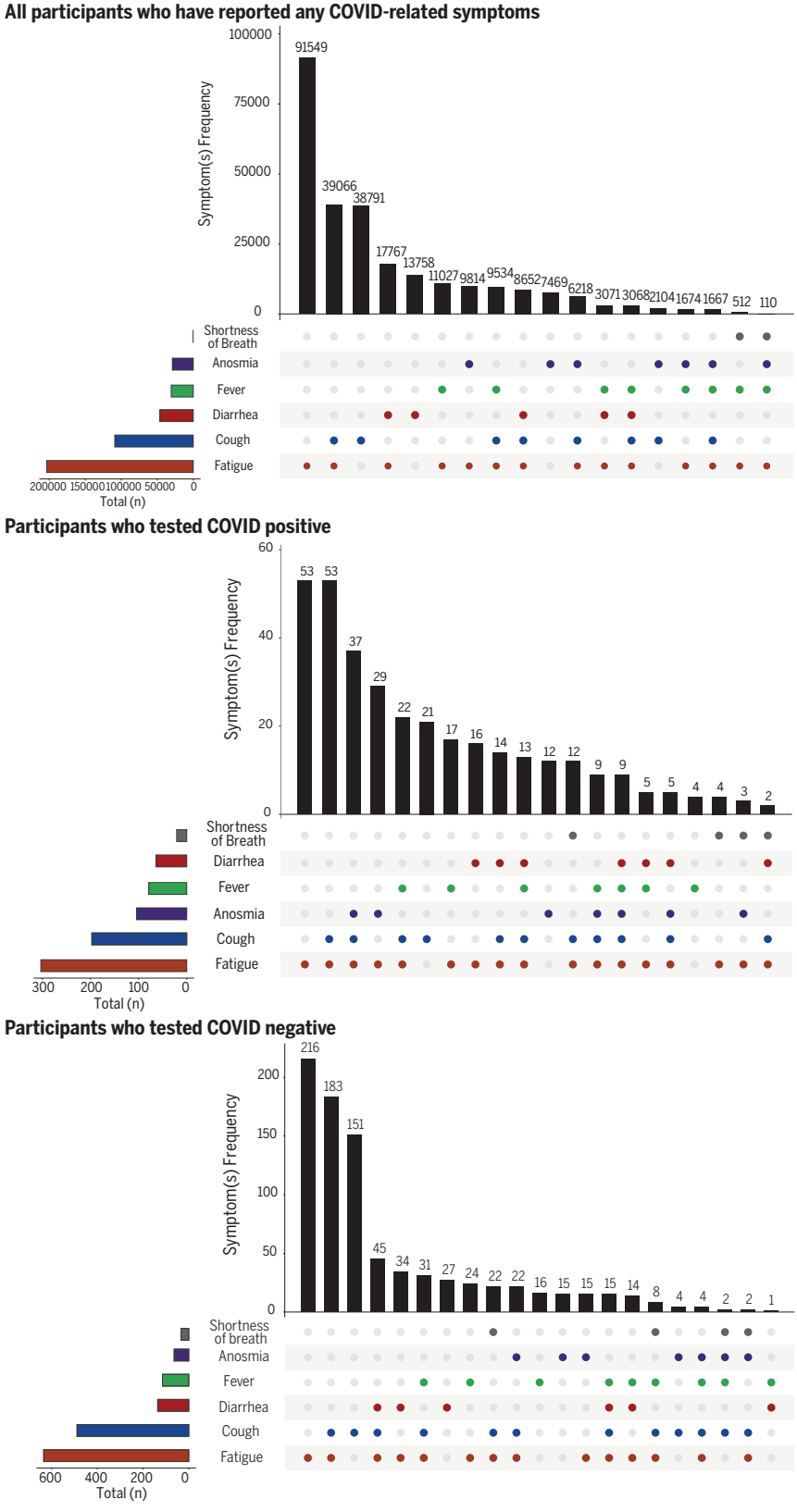


Fig. 3. Symptoms reported through the COVID Symptom Study app. By 27 March 2020, 265,851 individuals in the United Kingdom reported any symptom potentially associated with COVID-19 (**top**). Participants provided data on whether they were tested for COVID-19, as well as the test result. 1176 individuals reported having received a COVID-19 test (0.4% of those with symptoms). Symptom frequencies among those who tested positive (**middle**; n = 340) versus negative (**bottom**; n = 836) are shown.

Symptom Tracker) mobile app with in-kind contributions from Zoe Global Ltd., a digital health care company, and academic scientists from Massachusetts General Hospital and King's College London. By leveraging the established digital backbone of an application used for personal nutrition studies, the COVID Symptom Study app was launched in the United Kingdom on 24 March 2020 and became available in the United States on 29 March 2020 (https://covid.joinzoe.com/us). The COPE Consortium is committed to the shared international pursuit of combating COVID-19 and has worked with scientific collaborators and thought leaders in real-time epidemiology to prioritize data harmonization and sharing as part of the Coronavirus Census Collective (*17*).

The COVID Symptom Study enables self-reporting of data related to COVID-19 exposure and infections (Fig. 1). On first use, the app queries location, age, and core health risk factors. Daily prompts query for updates on interim symptoms, health care visits, and COVID-19 testing results. For those self-quarantining or seeking health care, the level of intervention and related outcomes are collected. Individuals without obvious symptoms are also encouraged to use the app. Through pushed software updates, we can add or modify questions in real time to test emerging hypotheses about COVID-19 symptoms and treatments. Notably, participants enrolled in ongoing epidemiologic studies, clinical cohorts, or clinical trials can provide informed consent to link survey data collected through the app to their preexisting study cohort data and any relevant biospecimens in a Health Insurance Portability and Accountability Act (HIPAA)–compliant and General Data Protection Regulation (GDPR)–compliant manner. A specific module is also provided for health care workers to determine the intensity and type of their direct patient care experiences, the availability and use of personal protective equipment (PPE), and work-related stress and anxiety.

Through rapid deployment of this tool, we can gain key insights into population dynamics of the disease (Fig. 2). By collecting participant-reported geospatial data, highlighted as a critical need for pandemic epidemiologic research (*15*), we can rapidly identify populations with highly prevalent symptoms in regions that may emerge as outbreak hotspots. An early snapshot of the first 1.6 million users in the United Kingdom over the first 5 days of use confirms the variability in symptoms reported across suspected COVID-19 cases and is useful for generating and testing broader hypotheses. At the time, users had a mean age of 41, ranged from 18 to 90 years old, and were 75% female. Graphic visualization of our initial results (Fig. 3) demonstrates that among those reporting symptoms by 27 March 2020 (n =

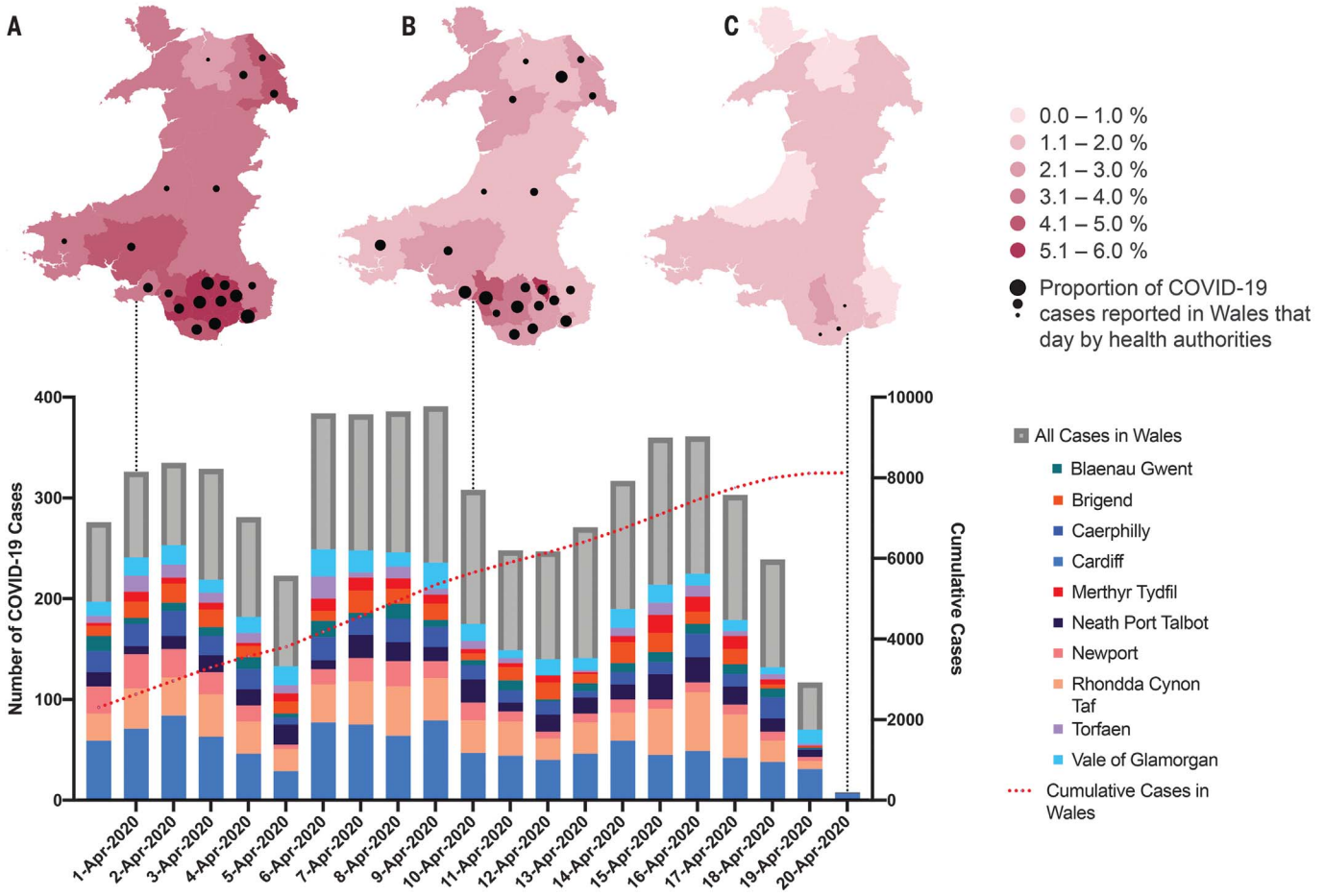


Fig. 4. Predicting COVID-19 cases on the basis of real-time symptom reporting in Wales, United Kingdom. This time series (bar graph) displays the number of new confirmed cases (gray bars) reported by the Public Health Wales NHS Trust between 31 March 2020 and 20 April 2020. After 2 April, case numbers appear to have declined through 5 April. However, our symptom-based prediction model (*18*), developed from symptom reports from untested users of the COVID Symptom Study app, showed a high proportion of predicted COVID-19 cases in southern Wales on 1 April [dark red areas in (**A**)]. Six days later, Welsh health authorities reported a subsequent peak in cases over a 4-day period (6 to 9 April), driven primarily by these southern regions (colored bars). By 10 April, new confirmed cases across Wales declined. However, on the basis of reported symptoms (**B**), regions in South Wales still had high predicted levels of COVID-19, which became apparent as a second spike in confirmed cases between 15 and 16 April. As of 20 April (**C**), predicted COVID-19 prevalence across Wales according to symptom reporting appears to be low, which corresponds to a flattening of the cumulative incidence curve. However, several regions in southern Wales still have relatively high reports of symptoms and appear at risk for subsequent cases of COVID-19. Black dots on the maps represent the relative proportion of positive tests reported by health authorities across Wales that day by region. The prediction mapping included data from 1,339,670 users of the COVID Symptom Study on 1 April; 998,244 users on 10 April; and 1,234,918 users on 20 April. Public Health Wales NHS Trust data were current as of 21 April 2020 at 13:00 local time, taken from the "Rapid COVID-19 Virology - Public" dashboard (accessed via https://phw.nhs.wales/), and downloaded on 22 April 2020 at 12:30 p.m. Eastern Standard Time.

265,851 individuals), the most common symptoms were fatigue and cough, followed by diarrhea, fever, and anosmia. Shortness of breath was reported relatively rarely. Only 0.4% (n = 1176) of individuals reporting possible COVID-19 symptoms reported receiving a quantitative polymerase chain reaction test for COVID-19.

A comparison of symptomatic users who reported receiving a test within the initial launch period generated several hypotheses for future study with the growing dataset. The frequency of cough or fatigue alone or in combination commonly led to testing but was not a particularly accurate predictor of a positive test. Similarly, no individuals who reported diarrhea in the absence of other symptoms tested positive. Notably, more complex presentations with cough and/or fatigue and at least one additional symptom, including less commonly appreciated complaints such as diarrhea and anosmia, appeared to be enriched among those with positive test results relative to those with negative results. In particular, anosmia may be a more predictive symptom, as it was more common than fever in individuals who tested positive. Indeed, in subsequent analyses with a larger sample set, we have shown that anosmia appears to be a strong predictor for COVID-19 (*18*). By contrast, fever alone was not particularly discriminatory. However, when fever was present in combination with less appreciated symptoms, a greater frequency of positive tests was observed. These findings suggest that perhaps individuals with complex or multiple-symptom (three or more) presentations should be prioritized for testing. Concerningly, 20% of individuals reported complex symptoms

(cough and/or fatigue plus at least one of anosmia, diarrhea, or fever) but had not yet been tested, representing a substantial population that appears to be at elevated risk for the disease. Additional work is warranted to confirm whether complex or multiple-symptom cases can accurately predict COVID-19 incidence.

Building on these initial findings, our team subsequently developed a weighted prediction model based on the symptoms of more than 2 million individual app users (*18*). By using this prediction model, we demonstrate the potential utility of the COVID Symptom Study app to collect data for long-term studies as well as for immediate public health planning. In southern Wales in the United Kingdom, users reported symptoms that predicted, 5 to 7 days in advance, two spikes in the number of confirmed positive COVID-19 cases reported by public health authorities (Fig. 4). Conversely, a decline in reports of symptoms preceded a drop in confirmed cases by several days. These results demonstrate that this app prospectively captures the dynamics of COVID-19 incidence days in advance of traditional measures, such as positive tests, hospitalizations, or mortality. We are currently planning additional studies using a broadly representative sample of individuals who will undergo uniform COVID-19 testing to further validate our approach to symptom-based modeling of incidence. These data demonstrate compelling evidence for the potential predictive power of our approach, which will improve as more data are collected to inform the model. Further, our data highlight the potential utility of real-time symptom tracking to help guide allocation of resources for testing and treatment as well as recommendations for lockdown or easement in specific areas.

With additional data collection, we will also apply big data approaches (e.g., machine learning) to identify emerging patterns in dynamic settings of exposure, onset of symptoms, disease trajectory, and clinical outcomes. Our launch of the app within several large epidemiology cohorts that have previously gathered longitudinal data on lifestyle, diet and health factors, and genetic information will allow investigation of a much broader range of putative risk factors for COVID-19 outcomes. With additional follow-up, we will also be positioned to investigate long-term effects of COVID-19, including mental health, disability, mortality, and financial outcomes. Mobile technology can also supplement recently launched clinical trials or biobanking protocols already embedded within clinical settings. In collaboration with the Stand Up to Cancer foundation, we have also developed a strategy to track information among individuals living with cancer, including those enrolled in clinical trials. At the Massachusetts General Hospital and Brigham and Women's Hospital, we are deploying the tool within several clinical studies, centralized biobanking efforts, and health care worker surveillance programs. Health care workers are particularly vulnerable to COVID-19's effects beyond infection, including work hazards from PPE shortages, emotional stress, and absenteeism. Real-time data generation focused within these populations will be critical to optimally allocate resources to protect our health care workforce and assess its efficacy.

Even so, our approach has limitations. We recognize that a smartphone application does not represent a random sampling of the population. However, this is an inherent limitation of any epidemiologic study that relies on voluntary participation. Our approach has the benefit of allowing rapid deployment across a large cross section of the population during a major public health crisis. With time and continued use, the large number of participants will include a sufficient quantity of users within key subgroups such that we can adjust our methodology for potential sources of confounding. By engaging cohorts with underrepresented populations, such as the BWHS in the United States, we also hope to leverage existing investigator-participant relationships to encourage enrollment of individuals from populations that have traditionally been challenging to recruit. Moreover, by encouraging longitudinal, prospective data collection, we can capture associations based on within-person variation over time, a notable advantage over repeated cross-sectional surveys that introduce considerable between-person variation. In the near future, we hope to release our app as fair-use open source software to facilitate translation and development in other regions. We have begun working with colleagues in Canada, Australia, and Sweden to implement this tool within their countries. We have also developed a practical toolkit to assist clinical researchers with local institutional review board and regulatory approval to facilitate deployment within research studies (www.monganinstitute.org/cope-consortium). This toolkit includes full details of the mobile app's questions, consent documents, privacy policies, and terms of use. With broader implementation, data generated from the COVID Symptom Study app are increasingly being linked to the public health response within the National Health Service (NHS) in the United Kingdom. The app is endorsed by the Welsh government, NHS Wales, the Scottish government, and NHS Scotland, and our scientific team updates the U.K. chief scientific officer daily. We are working to develop a similar approach in the United States. However, the lack of a national health care system has required a strategy focused on engaging local public health leaders. For example, we have partnered with the University of Texas School of Public Health to conduct statewide surveillance to support public health decision-making, especially as the Texas state government begins softening mitigation strategies.

Our approach demonstrates a proof of concept for rapid repurposing of existing data collection methods to implement scalable real-time collection of population-level data during a fast-moving global health crisis. We call on our colleagues to work with us so that we may deploy all of the tools at our disposal to address this unprecedented public health challenge.

REFERENCES AND NOTES

1. M. Lipsitch, D. L. Swerdlow, L. Finelli, *N. Engl. J. Med.* **382**, 1194–1196 (2020).
2. G. A. FitzGerald, *Science* **367**, 1434 (2020).
3. M. Day, *BMJ* **368**, m1086 (2020).
4. L. Fang, G. Karakiulakis, M. Roth, *Lancet Respir. Med.* **8**, e21 (2020).
5. W. J. Guan *et al*., *N. Engl. J. Med.* **382**, 1708–1720 (2020).
6. D. Wang *et al*., *JAMA* **323**, 1061 (2020).
7. C. Wu *et al*., *JAMA Intern. Med.* 10.1001/jamainternmed.2020.0994 (2020).
8. X. Yang *et al*., *Lancet Respir. Med.* 10.1016/S2213-2600(20)30079-5 (2020).
9. F. Zhou *et al*., *Lancet* **395**, 1054–1062 (2020).
10. N. Chen *et al*., *Lancet* **395**, 507–513 (2020).
11. C. Huang *et al*., *Lancet* **395**, 497–506 (2020).
12. L. Pan *et al*., *Am. J. Gastroenterol.* **115**, 766–773 (2020).
13. X. W. Xu *et al*., *BMJ* **368**, m606 (2020).
14. J. S. Brownstein, C. C. Freifeld, L. C. Madoff, *N. Engl. J. Med.* **360**, 2153–2157 (2009).
15. B. Xu, M. U. G. Kraemer; Open COVID-19 Data Curation Group, *Lancet Infect. Dis.* **20**, 534 (2020).
16. P. J. Lyons, "Coronavirus Briefing: What Happened Today," *The New York Times* (30 March 2020).
17. E. Segal *et al*., *Nat. Med.* 10.1038/s41591-020-0929-x (2020).
18. C. Menni *et al*., *Nat. Med.* 10.1038/s41591-020-0916-2 (2020).
19. D. A. Drew *et al*., medRxiv [Preprint]. 6 April 2020; https://doi.org/10.1101/2020.04.02.20051334.
20. MGHcteu, MGHcteu/ScienceCOPEMethodsCode: Science v1.0.1, Version 1.0.1, Zenodo (2020); https://doi.org/10.5281/zenodo.3765955.

ACKNOWLEDGMENTS

Use of the COVID-19 app in cohorts and informed consent as described was approved by the Partners Human Research Committee (Protocol 2020P000909) and is registered on ClinicalTrials.gov as NCT04331509. We acknowledge J. Capdevila Pujol of Zoe Global Ltd. for his contributions to this work. **Funding:** Zoe Global Ltd. provided in kind support for all aspects of building, running, and supporting the tracking app and service to all users worldwide. C.J.S., C.M., M.F., T.V., C.H.S., M.J.C., S.O., and T.D.S. were supported by the Wellcome Trust and EPSRC (WT212904/Z/18/Z, WT203148/Z/16/Z, and WT213038/Z/18/Z), the NIHR GSTT/KCL Biomedical Research Centre, MRC/BHF (MR/M016560/1), the NIHR, and the Alzheimer's Society (AS-JF-17-011). A.T.C. is the Stuart and Suzanne Steele MGH Research Scholar and Stand Up to Cancer scientist. L.H.N., D.A.D., and A.T.C. were supported by the Massachusetts Consortium on Pathogen Readiness (MassCPR), M. Schwartz, and L. Schwartz. **Author contributions:** D.A.D. and L.H.N. drafted the manuscript. All authors contributed to the conceptualization, methodology, formal analysis, investigation, resources, data curation, and review and editing of the manuscript. A.T.C. and T.D.S. supervised the study and acquired funding. **Competing interests:** T.D.S. is a consultant for Zoe Global Ltd. J.W. is an employee of Zoe Global Ltd. D.A.D. and A.T.C. previously served as investigators in a clinical trial of diet and lifestyle using a

distinct mobile application that was supported by Zoe Global Ltd. The other authors have no conflicts of interest to declare. **Data and materials availability:** This manuscript was previously made available as a preprint on medRxiv.org (*19*). Code used to generate maps is published and available in Zenodo (*20*). Data collected in the app are being shared with other health researchers through the NHS-funded Health Data Research UK (HDRUK)/SAIL Consortium, housed in the UK Secure e-Research Platform (UKSeRP) in Swansea, Wales. Anonymized data are available to be shared with bona fide researchers through HDRUK according to their protocols in the public interest. See https://healthdatagateway.org/detail/9b604483-9cdc-41b2-b82c-14ee3dd705f6. U.S. investigators are encouraged to coordinate data requests through the COPE Consortium (www.monganinstitute.org/cope-consortium). Data updates can be found on https://covid.joinzoe.com. For aggregate deidentified "snapshot" datasets used for analyses provided here, please request time-stamped datasets by email to predict@mgh.harvard.edu for data files, "COVIDSymptomTrackerData_03292020" (Figs. 2 and 3), or "COVIDSymptomTrackerData_04/21/20" (Fig. 4).

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/368/6497/1362/suppl/DC1
Figs. S1 to S9
COPE Consortium Members
MDAR Reproducibility Checklist

2 April 2020; accepted 30 April 2020
Published online 5 May 2020
10.1126/science.abc0473

MEASLES

Measles virus and rinderpest virus divergence dated to the sixth century BCE

Ariane Düx[1,2]*, Sebastian Lequime[3]*, Livia Victoria Patrono[1,2], Bram Vrancken[3], Sengül Boral[4], Jan F. Gogarten[1,2], Antonia Hilbig[1], David Horst[4], Kevin Merkel[1,2], Baptiste Prepoint[2,5], Sabine Santibanez[6], Jasmin Schlotterbeck[2], Marc A. Suchard[7,8,9], Markus Ulrich[1], Navena Widulin[10], Annette Mankertz[6], Fabian H. Leendertz[1], Kyle Harper[11], Thomas Schnalke[10], Philippe Lemey[3]†, Sébastien Calvignac-Spencer[1,2]†‡

Many infectious diseases are thought to have emerged in humans after the Neolithic revolution. Although it is broadly accepted that this also applies to measles, the exact date of emergence for this disease is controversial. We sequenced the genome of a 1912 measles virus and used selection-aware molecular clock modeling to determine the divergence date of measles virus and rinderpest virus. This divergence date represents the earliest possible date for the establishment of measles in human populations. Our analyses show that the measles virus potentially arose as early as the sixth century BCE, possibly coinciding with the rise of large cities.

Measles is a highly contagious viral disease that presents with rash, fever, and respiratory symptoms. Before a live-attenuated vaccine was developed in the 1960s, the disease affected the vast majority of children (*1*, *2*). Global vaccination campaigns resulted in a marked reduction of measles transmission and fatal cases, and the World Health Organization (WHO) has proclaimed an elimination goal. However, the disease still caused an estimated 110,000 deaths in 2017 (*3*), and incidence has recently been on the rise (*4*). Measles is caused by *Measles morbillivirus* (MeV), a negative-sense single-stranded RNA virus from the family *Paramyxoviridae* (Order: *Mononegavirales*). MeV is an exclusively human pathogen whose closest relative was the now eradicated *Rinderpest morbillivirus* (RPV), a devastating cattle pathogen (*5*). It is generally accepted that measles emergence resulted from a spillover from cattle to humans, although the directionality of this cross-species transmission event has never been formally established (supplementary text S1) (*6*).

It is unclear when measles first became endemic in human populations, but assuming an origin in cattle, the earliest possible date of MeV emergence is defined by the MeV-RPV divergence time. Several studies have provided estimates for this date using molecular-clock analyses (*7*–*10*), with the most reliable (and oldest) estimate falling at the end of the ninth century CE {mean, 899 CE [95% highest posterior density (HPD) interval, 597 to 1144 CE]} (*8*). We reassessed the MeV-RPV divergence time using advanced, selection-aware Bayesian molecular-clock modeling (*11*) on a dataset of heterochronous MeV genomes, including the oldest human RNA virus genome sequenced to date, and show that a considerably earlier emergence can no longer be excluded.

Our reexamination was prompted by the broadly accepted view that molecular dating based on tip date calibration—the method used in previous efforts to estimate the timing of MeV-RPV divergence—underestimates deep divergence times (*8*). Rapid short-term substitution rates captured with tip date calibration often cannot be applied over long evolutionary time scales because of the effects of long-term purifying selection and substitution saturation. This causes a discrepancy between short- and long-term substitution rates, which is referred to as the time-dependent rate phenomenon (*12*, *13*). Because measurement time scales matter, a first step to arrive at accurate estimates is to maximize the time depth of tip date calibration—for example, through the use of ancient viral sequences (*14*, *15*).

RNA tends to be much less stable in the environment than DNA, making the recovery of MeV genetic material from archaeological remains unlikely (*16*). Pathology collections represent a more realistic source of MeV sequences that predate the oldest MeV genome: the genome of the Edmonston strain that was isolated in 1954 and attenuated to become the first measles vaccine. We examined a collection of lung specimens gathered by Rudolf Virchow and his successors between the 1870s and 1930s and preserved by the Berlin Museum of Medical History at the Charité (Berlin, Germany) and identified a 1912 case diagnosed with fatal measles-related bronchopneumonia (Fig. 1, fig. S1, and supplementary text S2 and S3). To retrieve MeV genetic material from this specimen, we first heat-treated 200 mg of the formalin-fixed lung tissue to reverse macromolecule cross-links induced by formalin and subsequently performed nucleic acid extraction (*17*). After deoxyribonuclease treatment and ribosomal RNA depletion, we built high-throughput sequencing libraries and shotgun sequenced them on Illumina platforms. We generated 27,328,219 high-quality reads, of which 0.46% were mapped to a MeV genome. Median insert size varied between 95 and 136 nucleotides (nt), and little damage was observed, suggesting good preservation of RNA molecules (fig. S2, table S1, and supplementary text S4). The resulting 10,960 unique MeV reads allowed us to reconstruct an almost complete 1912 MeV genome: 15,257 of the 15,894 nt in the MeV strain Edmonston (AF266288) were covered by at least three unique reads (11,988 nt by at least 20 reads; mean coverage 54x).

In addition to the 1912 genome and the 1954 Edmonston genome, only two genomes have been determined from MeV isolated before 1990 (Mvi/Lyon.FRA/77, HM562899; and T11wild, AB481087). We therefore searched the strain collection of the German National Reference Laboratory (Robert Koch Institute, Berlin, Germany) for pre-1990 isolates. We found two strains from the pre-vaccine era isolated in 1960 by the National Reference Laboratory of former Czechoslovakia in Prague (MVi/Prague.CZE/60/1 and MVi/Prague.CZE/60/2) (*18*). We performed serial passages of these strains and determined their genome sequences at a mean coverage of 109x and 70x, respectively. The two genomes were nearly identical, differing at only four sites.

We performed Bayesian and maximum likelihood (ML) phylogenetic analyses to investigate the phylogenetic placement of the 1912 and 1960 genomes with respect to 127 available MeV genomes. Tip-dated Bayesian phylogenetic trees placed the 1912 genome as a sister lineage to all modern genomes, whereas the two genomes from 1960 clustered together with the Edmonston strain (genotype A) (fig. S3). The placement of the 1912 genome in the dated-tip tree was consistent with its placement in a non-clock ML tree reconstruction and with the rooting of a dated-tip tree excluding the 1912 genome (fig. S4 A and B). The relatedness of the 1912 and 1960 genomes to now extinct MeV lineages is in line with a

[1]Epidemiology of Highly Pathogenic Microorganisms Project Group, Robert Koch Institute, Berlin, Germany. [2]Viral Evolution Project Group, Robert Koch Institute, Berlin, Germany. [3]Laboratory of Clinical and Evolutionary Virology, Department of Microbiology, Immunology and Transplantation, Rega Institute, Katholieke Universiteit (KU) Leuven, Leuven, Belgium. [4]Institute for Pathology, Charité, Berlin, Germany. [5]Département de Biologie, Ecole Normale Supérieure, PSL Université Paris, Paris, France. [6]National Reference Centre for Measles, Mumps, and Rubella, Robert Koch Institute, Berlin, Germany. [7]Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, USA. [8]Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [9]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [10]Berlin Museum of Medical History, Charité, Berlin, Germany. [11]Department of Classics and Letters, University of Oklahoma, Norman, OK, USA.
*These authors contributed equally to this work. †These authors contributed equally to this work.
‡Corresponding author. Email: calvignacs@rki.de

Fig. 1. Formalin-fixed lung specimen collected in 1912 in Berlin. Specimen is from a 2-year-old girl diagnosed with measles-related bronchopneumonia (museum object ID: BMM 655/1912).

marked reduction of MeV genetic diversity during the 20th century as a product of massive vaccination efforts.

Having extended the depth of MeV tip date calibration, we subsequently focused our attention on estimating the timing of MeV-RPV divergence. We assembled a dataset of 51 genomes comprising MeV (including one of the 1960 genomes and the 1912 genome), RPV, and Peste des petits ruminants virus (PPRV; the closest relative to MeV-RPV) sequences, ensuring that they represented the known genetic diversity of these viruses (table S2). Before inferring a time-scaled evolutionary history for this dataset by using a Bayesian phylogenetic framework, we assessed its temporal signal and tested it for substitution saturation. We confirmed a strong temporal signal (fig. S5 and table S3) and did not identify strong substitution saturation (table S4). We constructed a series of increasingly complex evolutionary models to accommodate various sources of rate heterogeneity. Models ranged from a standard codon substitution model with a strict molecular clock assumption to a codon substitution model with time-varying selection combined with a clade-specific rate for PPRV and additional branch-specific random effects on the substitution rate. Adequately accommodating different sources of rate heterogeneity is known to provide a better correction for multiple hits in genetic distance estimation,

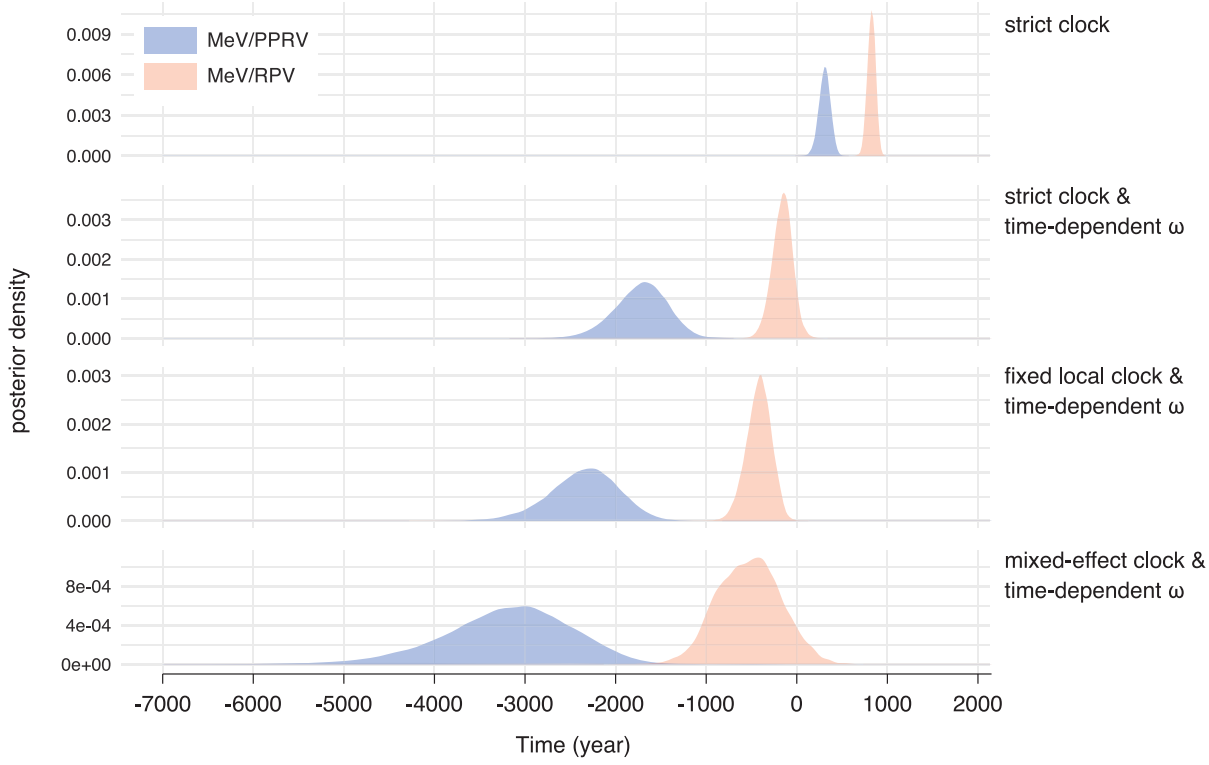


Fig. 2. Divergence time estimates for MeV and RPV (red) and for MeV-RPV and PPRV (blue) under increasingly complex evolutionary models. Estimates for parameters of interest (posterior mean and 95% highest posterior density interval) under each model are provided in table S5.

and the potential of codon substitution modeling in recovering deep viral divergence has been demonstrated specifically (*8*). This was reflected in the increasingly older estimates of MeV-RPV and PPRV-MeV-RPV divergence times and wider credible intervals for increasingly complex models (Fig. 2 and table S5). Parameter estimates of the substitution and clock models also provided evidence for a significant contribution of these different sources of rate heterogeneity to model fit improvement (table S5). We found a significantly negative coefficient for the time-dependent nonsynonymous/synonymous substitution rate ratio (ω) (*11*), indicating strong long-term purifying selection; a significantly positive coefficient for the fixed effect on the PPRV rate, indicating a faster evolutionary rate in this clade (as suggested by temporal signal analyses, fig. S5); and significant additional unexplained variation, as modeled by the random effects (table S5). Our most complex model therefore provided the best description of the evolutionary process and pushed back the divergence date of MeV and RPV, with a mean estimate at 528 BCE (95% HPD interval, 1174 BCE to 165 CE) (Fig. 3). These estimates were robust to (i) including or excluding the 1912 genome in the analyses (table S6), (ii) using a more conservative consensus genome for the 1912 sample (table S7), (iii) the prior specification on the age of the RPV genome or the inclusion of an additional RPV genome (table S7), and (iv) the coalescent prior specification (table S7). A comparison of the four models in their ability to recover the age of the 1912 genome indicated that the most complex model yielded the best estimate [1929 CE (95% HPD interval, 1889 to 1961 CE)] (fig. S6).

The MeV-RPV divergence time provides the earliest possible date for measles emergence in humans, which is now compatible with the emergence of this disease more than 2500 years ago. It seems plausible that the divergence of these lineages was closely followed by the cattle-to-human host jump and subsequent evolution into two distinct pathogens. However, the spillover could have occurred at any time between the MeV-RPV divergence and the time to the most recent common ancestor of all MeV known to infect humans [1880 CE (95% HPD interval, 1865 to 1893 CE)]. This raises the question of whether other sources of information can narrow down this time frame and agree with an earlier timing of measles emergence.

The earliest clear clinical description of measles is often attributed to the Persian physician Rhazes, writing in the 10th century CE (*19*). But Rhazes was extremely familiar with all available medical literature at his time and made use of earlier sources. Indian medical texts possibly describe measles several centuries before Rhazes (*20*). Although clear descriptions of measles are missing in the Hippocratic corpus and the Greek medical tradition (at least through the prolific second-century writer Galen), such absence alone cannot be decisive. Retrospective diagnosis from premodern medical texts is notoriously fraught, especially for diseases such as measles, whose symptoms were easily confused with a variety of other conditions. Measles differential diagnosis remained a challenge well into more recent times (*21*). Therefore, any number of the large-scale "pestilences" described in ancient sources from Europe or China could reflect MeV outbreaks.

An ancient origin of measles seems all the more plausible in the light of demographic changes that are compatible with our understanding of (contemporary) MeV epidemiology. Populations large enough to support continuous MeV transmission—larger than the MeV critical community size (CCS) of 250,000 to 500,000 individuals (*22*–*24*)—could not exist in Neolithic, Bronze Age, and early Iron Age settlements, which lacked both economic and political means to allow such numbers. Even if connectivity between such settlements may have created a larger pool of susceptible individuals, epidemiologists have held that given the speed with which measles epidemics occur and the efficacy of acquired immunity, MeV could not have become endemic in urban populations below the CCS (*25*). In the late first millennium BCE, technologies (both economic and political) crossed a threshold, promoting an upsurge in population sizes in Eurasia and South and East Asia. Although considerable uncertainty exists around population size estimates derived from ancient documents (for

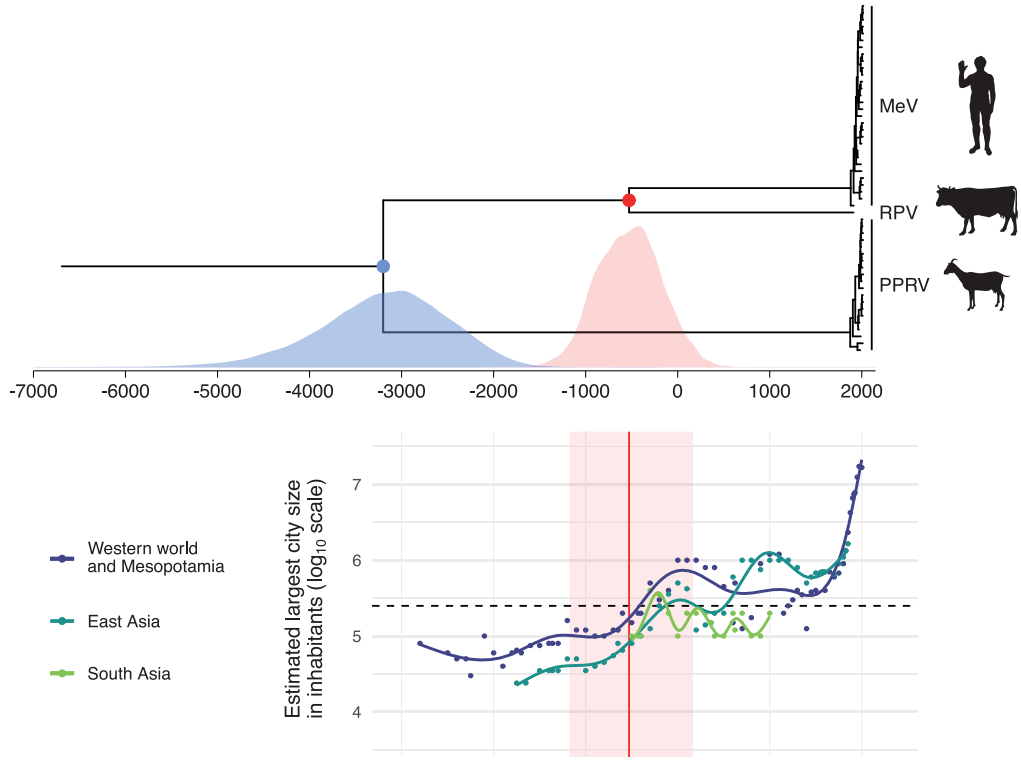


Fig. 3. Time-measured evolutionary history for MeV, RPV, and PPRV and largest city size over time in three well-studied regions of the world. (**Top**) Maximum clade credibility tree summarized from a Bayesian time-measured inference by using tip-dating and accounting for long-term purifying selection. The red and blue points represent the mean estimates for the divergence times between MeV and RPV then MeV-RPV and PPRV, respectively; the corresponding divergence date estimates are depicted below as marginal posterior distributions. (**Bottom**) The estimated size (log10 scale) of the largest city in the western world including Mesopotamia (dark blue), East Asia (dark green), and South Asia (green) over time. The red vertical line indicates the mean divergence time estimate between MeV and RPV, and the red area indicates its 95% highest posterior density interval. The dashed horizontal line represents the classical threshold for MeV maintenance in a population (250,000 individuals). Dots show data points according to Morris (*34*) and Inoue *et al*. (*26*). Each line represents the fit of a generalized additive model with a cubic spline smoothing function.

example, literary observations, travelers' reports, censuses, or references to the amount of food distributed in a city) or archaeological proxies (such as the size of city walls or a built-up area of settlement), there is broad agreement that a number of settlements in North Africa, India, China, Europe, and the Near East began to surpass the CCS for MeV by around 300 BCE, presumably for the first time in human history (Fig. 3) (*26*). From this period onward, there were consistently urban populations above the CCS for MeV.

On the basis of these considerations, our substantially older MeV-RPV divergence estimate provides grounds for sketching a new model of MeV's evolutionary history. Under this scenario, a bovine virus, the common ancestor of modern strains of RPV and MeV, circulated in large populations of cattle (and possibly wild ungulates) since its divergence from PPRV around the fourth millennium BCE [3199 BCE (95% HPD interval, 4632 to 1900 BCE)] (Fig. 3). As a fast-evolving RNA virus, it may have produced variants that were able to cross the species barrier on several occasions, but small human populations could only serve as dead-end hosts. Then, almost as soon as contiguous settlements reached sufficient sizes to maintain the virus' continuous transmission (Fig. 3), it emerged as a human pathogen, the progenitor of modern-day MeV. It has been suggested that numerous concurrent human-bovine epidemics in the early medieval period (here, 6th to 10th centuries CE) were caused by an immediate ancestor of MeV and RPV that was pathogenic to both cattle and humans (*27*). The new MeV-RPV divergence date allows for the same inference to be made for earlier concurrent human-bovine mortality events well attested in, for example, Roman sources from the fifth century BCE on (*28*). During the following centuries, introduction of MeV into naive human populations and/or flare-ups of the disease might have caused some ancient epidemics whose etiology remains uncertain.

Although our findings shed new light on the origin of measles, formally proving that the virus emerged soon after its divergence from RPV would require archaeological genomic evidence. Most studies on ancient viruses have thus far focused on viruses with a double-stranded DNA genome (*14*, *29*–*32*). However, genetic material of parvovirus B19 was also detected in early Neolithic skeletal remains, despite the relatively unstable nature of its single-stranded DNA genome (*15*). It remains to be determined whether viral RNA recovery from such ancient specimens is feasible. Recently, RNA was extracted from the remains of a 14,300-year-old Pleistocene canid preserved in permafrost (*33*). Although the majority of RNA fragments were extremely short (<30 nt), the authenticity of the sequences could be validated (*33*). Such advances highlight that it may not be completely impossible for ancient remains to still contain MeV RNA, especially if preserved under favorable circumstances, including natural mummification or preservation in cold environments (*16*). While awaiting such direct evidence, we believe that the proposed model of MeV evolution constitutes a compelling working hypothesis.

REFERENCES AND NOTES

1. A. D. Langmuir, *Am. J. Dis. Child.* **103**, 224–226 (1962).
2. W. J. Moss, *Lancet* **390**, 2490–2502 (2017).
3. A. Dabbagh *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **67**, 1323 (2018).
4. WHO, Provisional data based on monthly data reported to WHO (Geneva) as of January 2020; www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en
5. P. Roeder, J. Mariner, R. Kock, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120139 (2013).
6. N. D. Wolfe, C. P. Dunavan, J. Diamond, *Nature* **447**, 279–283 (2007).
7. Y. Furuse, A. Suzuki, H. Oshitani, *Virol. J.* **7**, 52 (2010).
8. J. O. Wertheim, S. L. Kosakovsky Pond, *Mol. Biol. Evol.* **28**, 3355–3365 (2011).
9. M. Muniraju *et al.*, *Emerg. Infect. Dis.* **20**, 2023–2033 (2014).
10. H. Kimura *et al.*, *Sci. Rep.* **5**, 11648 (2015).
11. J. V. Membrebe, M. A. Suchard, A. Rambaut, G. Baele, P. Lemey, *Mol. Biol. Evol.* **36**, 1793–1803 (2019).
12. S. Y. Ho, S. Duchêne, M. Molak, B. Shapiro, *Mol. Ecol.* **24**, 6007–6012 (2015).
13. P. Aiewsakun, A. Katzourakis, *J. Virol.* **90**, 7184–7195 (2016).
14. B. Mühlemann *et al.*, *Nature* **557**, 418–423 (2018).
15. B. Mühlemann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7557–7562 (2018).
16. O. Smith, M. T. Gilbert, in *Paleogenomics: Genome-Scale Analysis of Ancient DNA*, C. Lindqvist, O. P. Rajora, Eds. (Springer, 2019).
17. M. T. Gilbert *et al.*, *PLOS ONE* **2**, e537 (2007).
18. S. Santibanez, A. Heider, E. Gerike, A. Agafonov, E. Schreier, *J. Med. Virol.* **58**, 313–320 (1999).
19. R. Kim-Farley, in *The Cambridge World History of Human Disease*, K. F. Kiple, Ed. (Cambridge Univ. Press, 1993).
20. K. R. L. Gupta (translator), *Madhava Nidana: Ayurvedic System of Pathology* (Sri Satguru Publications, 1987).
21. B. A. Cunha, *Infect. Dis. Clin. North Am.* **18**, 79–100 (2004).
22. M. S. Bartlett, *J. R. Stat. Soc. [Ser A]* **120**, 48–70 (1957).
23. F. L. Black, *J. Theor. Biol.* **11**, 207–211 (1966).
24. M. J. Keeling, B. T. Grenfell, *Science* **275**, 65–67 (1997).
25. A. D. Cliff, P. Haggett, M. Smallman-Raynor, *Measles: An Historical Geography of a Major Human Viral Disease: From Global Expansion to Local Retreat, 1840–1990* (Blackwell, 1993).
26. H. Inoue *et al.*, *Soc. Sci. Hist.* **39**, 175–200 (2015).
27. T. P. Newfield, *J. Interdiscip. Hist.* **46**, 1–38 (2015).
28. C. A. Spinage, *Cattle Plague: A History* (Springer, 2003).
29. P. Biagini *et al.*, *N. Engl. J. Med.* **367**, 2057–2059 (2012).
30. A. T. Duggan *et al.*, *Curr. Biol.* **26**, 3407–3412 (2016).
31. Z. Patterson Ross *et al.*, *PLOS Pathog.* **14**, e1006750 (2018).
32. B. Krause-Kyora *et al.*, *eLife* **7**, e36666 (2018).
33. O. Smith *et al.*, *PLOS Biol.* **17**, e3000166 (2019).
34. I. Morris, *The Measure of Civilization: How Social Development Decides the Fate of Nations* (Princeton Univ. Press, 2013).
35. S. Lequime, slequime/measles-history: Publication. Zenodo (2020); doi: 10.5281/zenodo.3764907.

ACKNOWLEDGMENTS

The 1912 MeV genome was generated from a formalin-fixed lung specimen (museum object ID: BMM 655/1912) from the collection of the Berlin Museum of Medical History at the Charité (Berlin, Germany). Ethics approval was obtained from the ethics committee of the Charité (Berlin, Germany) under the reference number EA4/212/19. We thank the National Institute of Public Health, Prague, Czech Republic, for providing us with the MeV strains MVi/Prague.CZE/60/1 and MVi/Prague.CZE/60/2 and O. Smith and J. Wertheim for helpful suggestions. The Titan V GPU used for this research was donated by the NVIDIA Corporation. **Funding:** The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement ~725422-ReservoirDOCS). P.L. acknowledges support by the Research Foundation–Flanders ('Fonds voor Wetenschappelijk Onderzoek–Vlaanderen,' FWO: G066215N, G0D5117N, and G0B9317N). S.L. and B.V. are postdoctoral research fellows funded by the FWO. M.A.S. was partially supported through National Institutes of Health grant U19 AI135995. **Author contributions:** Conceptualization: P.L. and S.C.-S.; data curation: A.D., S.L., B.V., M.A.S., P.L., and S.C.-S.; formal analysis: A.D., S.L., B.V., J.F.G., M.U., P.L., and S.C.-S.; funding acquisition: F.H.L. and S.C.-S.; investigation: A.D., L.V.P., S.B., A.H., D.H., K.M., B.P., and J.S.; methodology: S.L., B.V., M.A.S., and P.L.; project administration: F.H.L., P.L., and S.C.-S.; resources: S.S., N.W., A.M., F.H.L., T.S., P.L., and S.C.-S.; software: S.L., B.V., M.A.S., and P.L.; supervision: P.L. and S.C.-S.; validation: A.D., M.U., J.F.G., and S.C.-S.; visualization: S.L.; writing, original draft: A.D., S.L., K.H., P.L., and S.C.-S.; writing, review and editing: all authors. **Competing interests:** All authors declare no competing interests. **Data and materials availability:** The sequencing data for this study have been deposited in the European Nucleotide Archive (ENA) at the European Molecular Biology Laboratory–European Bioinformatics Institute under accession no. PRJEB36265 (www.ebi.ac.uk/ena/data/view/PRJEB36265). Human reads have been removed from 1912 sequencing files before uploading (supplementary materials). Alignments, trees, and BEAST xml files are available at https://github.com/slequime/measles-history (*35*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1367/suppl/DC1
Materials and Methods
Supplementary Text S1 to S4
Figs. S1 to S6
Tables S1 to S7
References (*36*–*83*)

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

17 January 2020; accepted 30 April 2020
10.1126/science.aba9411

IMMUNOMETABOLISM

T cells with dysfunctional mitochondria induce multimorbidity and premature senescence

Gabriela Desdín-Micó[1,2], Gonzalo Soto-Heredero[1,2]*, Juan Francisco Aranda[1,2]*, Jorge Oller[1,2]*, Elisa Carrasco[1,2], Enrique Gabandé-Rodríguez[1,2], Eva Maria Blanco[1,2], Arantzazu Alfranca[3], Lorena Cussó[4,5,6,7], Manuel Desco[4,5,6,7], Borja Ibañez[5,8,9], Arancha R. Gortazar[10], Pablo Fernández-Marcos[11], Maria N. Navarro[2,3], Bruno Hernaez[2], Antonio Alcamí[2], Francesc Baixauli[12]†, María Mittelbrunn[1,2]†‡

The effect of immunometabolism on age-associated diseases remains uncertain. In this work, we show that T cells with dysfunctional mitochondria owing to mitochondrial transcription factor A (TFAM) deficiency act as accelerators of senescence. In mice, these cells instigate multiple aging-related features, including metabolic, cognitive, physical, and cardiovascular alterations, which together result in premature death. T cell metabolic failure induces the accumulation of circulating cytokines, which resembles the chronic inflammation that is characteristic of aging ("inflammaging"). This cytokine storm itself acts as a systemic inducer of senescence. Blocking tumor necrosis factor–α signaling or preventing senescence with nicotinamide adenine dinucleotide precursors partially rescues premature aging in mice with *Tfam*-deficient T cells. Thus, T cells can regulate organismal fitness and life span, which highlights the importance of tight immunometabolic control in both aging and the onset of age-associated diseases.

The emerging field of immunometabolism (*1*–*5*) has expanded therapeutic opportunities for the treatment of not only inflammatory and autoimmune disorders (*6*) but also metabolic diseases and cancer (*7*–*10*). However, the utility of targeting immunometabolism to prevent the onset of age-associated diseases and multimorbidity has not been previously addressed.

An age-related decline in mitochondrial function has been observed in various cells and tissues, including T cells (*11*). To investigate the consequences of T cell metabolic decline in healthy aging, we used *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice. *Tfam* is a nuclear gene that encodes mitochondrial transcription factor A (TFAM), which both stabilizes mitochondrial DNA (mtDNA) and initiates mtDNA replication (*12*). In *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice, *Tfam* is depleted in both CD4 and CD8 T lymphocytes (*13*). T cell *Tfam* deficiency reduced the numbers of total circulating CD4 and CD8 T cells (fig. S1A). Lack of *Tfam* resulted in a sharp decrease in T cell mtDNA content (fig. S1B) and the failure to express key electron transport chain components, which reprogrammed T cell metabolism toward glycolysis (*13*) (Fig. 1, A to C). T cells from young (2-month-old) *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice recapitulated features of the mitochondrial dysfunction that appears in aged (22-month-old) wild-type mice (Fig. 1, A to C). This mitochondrial decline was associated with T helper 1 (T_H1) cell skewing, characterized by higher secretion of inflammatory type 1 cytokines interferon-γ (IFN-γ) and tumor necrosis factor–α (TNF-α), and the increased expression of the T_H1 cell master regulator T-bet (Fig. 1, D and E, and fig. S1C). In addition to this proinflammatory phenotype, young (2-month-old) *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice were immunocompromised to a similar extent as old (22-month-old) wild-type mice. We infected *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice with ectromelia virus (ECTV), a highly virulent mouse poxvirus that causes a disease similar to human smallpox, and compared them with young control or old wild-type mice. Infection with ECTV killed old wild-type mice and young *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice within the first 10 days of infection, whereas all young controls were able to survive this acute infection (Fig. 1F). Thus, *Tfam*-deficient T cells recapitulate the metabolic, phenotypic, and functional features of aged T cells. Seven-month-old *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice exhibited premature inflammaging, with circulating levels of the inflammaging-associated cytokines IL-6, IFN-γ, and TNF-α similar to the levels observed in 22-month-old wild-type mice (Fig. 1G and fig. S1, D and E). In humans, inflammaging predicts susceptibility to cardiovascular diseases, neurodegeneration, frailty, and multimorbidity (*14*–*16*). *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice displayed a prematurely aged appearance from the age of 7 months (Fig. 1H), which progressed to anemia, kyphosis, and low body weight (Fig. 1, I to K, and fig. S1F). An additional indicator of aging in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice was the significant thinning of hypodermal fat (Fig. 1L). Consistent with a premature aging phenotype, metabolic cage experiments revealed that *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice were less active and slower than controls, despite higher energy expenditure (Fig. 1M and fig. S1G). Notably, mean life span in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice was half that of controls (483 days versus 984 days) (Fig. 1N). Although *Tfam* deletion in regulatory T cells induces lethal autoimmunity (*17*), *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice showed no differences in serum autoantibody levels (fig. S1H).

To evaluate the potential of *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice as a model of age-related multimorbidity, we analyzed muscular, cardiovascular, and cognitive fitness. Histological analyses of *Tfam*$^{fl/fl}$ *Cd4*^{Cre} muscle tissue revealed a reduced fiber diameter (Fig. 2A). Imaging analysis of the gastrocnemius muscle after injection of ^{18}F-fluorodeoxyglucose (^{18}F-FDG) revealed significantly lower glucose uptake in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (Fig. 2B), and lower skeletal muscle strength was confirmed by the grip strength test (Fig. 2C). Moreover, *Tfam*$^{fl/fl}$ *Cd4*^{Cre} muscle up-regulated the expression of genes encoding the ubiquitin ligases MuRF-1 and Atrogin-1 as well as the expression of inflammatory markers (Fig. 2D and fig. S2A). *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice displayed a marked loss of gonadal white adipose tissue (gWAT) mass and smaller adipocytes (Fig. 2, E and F). These differences were accompanied by elevated expression levels of the adipose triglyceride lipase in gWAT and elevated circulating levels of nonesterified fatty acids, both of which were indicative of lipolysis induction (Fig. 2, G and H). Thus, the loss of T cell immunometabolic control induces sarcopenia and lipolysis. Additionally, *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice showed evidence of cardiac atrophy (Fig. 2I and fig. S2B). Histological and echocardiographic analyses revealed relative reduction in left ventricular thickness, reduced left ventricular diameter and volume, and smaller cardiomyocyte size (Fig. 2J and fig. S2, C to F). These features were accompanied by a higher heart rate (Fig. 2K) and the induction of cardiac stress markers *Foxo3a* and *Nppa* (fig. S2G). *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice displayed diastolic failure, characterized by reduced cardiac output, as well as elevated normalized lung weight with signs of lung congestion, as shown by computed tomography scan (Fig. 2, L and M, and fig. S2H).

[1]Instituto de Investigación Sanitaria Hospital 12 de Octubre (imas12), Madrid, Spain. [2]Departamento de Biología Molecular, Centro de Biología Molecular Severo Ochoa (CBMSO), Consejo Superior de Investigaciones Científicas (CSIC)-Universidad Autónoma de Madrid (UAM), Madrid, Spain. [3]Hospital Universitario de la Princesa, Madrid, Spain. [4]Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Madrid, Spain. [5]Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain. [6]Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain. [7]Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, Spain. [8]IIS-Fundación Jiménez Díaz, Madrid, Spain. [9]Centro de Investigación Biomédica en Red, Enfermedades Cardiovasculares (CIBERCV), Madrid, Spain. [10]Bone Physiopathology Laboratory, Applied Molecular Medicine Institute (IMMA), Universidad San Pablo-CEU, Madrid, Spain. [11]Metabolic Syndrome Group - BIOPROMET, Madrid Institute for Advanced Studies - IMDEA Food, CEI UAM +CSIC, Madrid, Spain. [12]Department of Immunometabolism, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: mmittelbrunn@cbm.csic.es

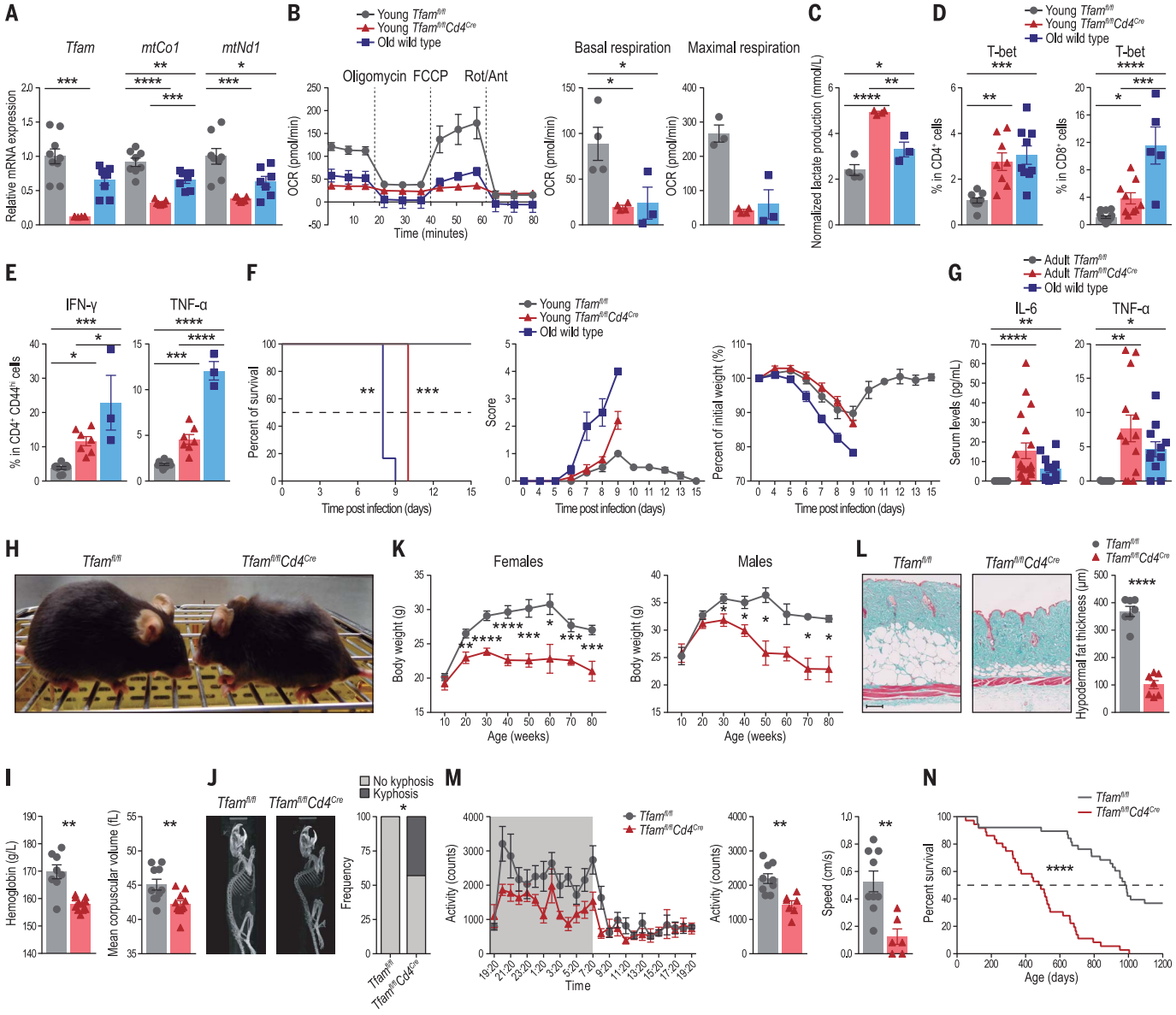


Fig. 1. Mitochondrial dysfunction in T cells causes premature aging. (**A**) *Tfam*, *mtCo1*, and *mtNd1* mRNA levels in peripheral blood CD4$^+$ T cells from young (2-month-old) *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice and old (22-month-old) wild-type (WT) mice (n = 6 to 9 mice per group). (**B**) (Left) Oxygen consumption rates (OCR) in activated CD4$^+$ T cells from *Tfam*$^{fl/fl}$, *Tfam*$^{fl/fl}$ *Cd4*^{Cre}, and old WT animals. Basal respiration (center) and maximal respiratory capacity (right) are shown (n = 3 to 4). FCCP, carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone; Rot, rotenone; Ant, antimycin A. (**C**) Lactate content in the supernatant of activated CD4$^+$ T cells (n = 3 to 4). (**D**) Percentages of CD4$^+$ (left) and CD8$^+$ (right) T cells positive for the T_H1 cell transcription factor T-bet (n = 5 to 14). (**E**) Percentages of CD4$^+$CD44^{hi} T cells staining positive for intracellular IFN-γ and TNF-α (n = 3 to 9). (**F**) Postinfection survival curves (left) for *Tfam*$^{fl/fl}$, *Tfam*$^{fl/fl}$ *Cd4*^{Cre}, and old WT mice inoculated subcutaneously with ECTV [10^3 plaque-forming units (PFU) per mouse]. ECTV-infected mice were monitored daily for clinical signs of illness (center) and change from initial body weight (right). Signs of illness are expressed as means ± SEM using an individual score ranging from 0 for healthy animals to 4 for severely diseased animals (n = 5 to 7). (**G**) Serum levels of inflammatory cytokines IL-6 and TNF-α detected by multiplex in 7-month-old *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice and 22-month-old WT mice (n = 10 to 19). (**H**) Representative photograph showing the deteriorated physical appearance of a *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mouse (right) compared with a control littermate (left), both aged 7 months. (**I**) Hematological parameters in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} and *Tfam*$^{fl/fl}$ mice (n = 8 to 11 5-month-old mice). (**J**) Quantification of spine curvature by computed tomography (CT) scans (left) and percentage of mice presenting lordokyphosis (right) in 5-month-old *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 7 to 8). (**K**) Body weight evolution in *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} female (left) and male (right) mice (n = 8 to 20). (**L**) Representative skin sections stained with Masson trichrome (left) and quantification of hypodermal fat thickness (right). At least 10 measurements were performed per animal. The graph shows mean values for n = 7 mice at 7 months of age. Scale bar, 100 μm. (**M**) Activity time course over a 24-hour cycle (left) and mean dark period activity (center) and speed (right) assessed in metabolic cages (n = 6 to 9 7-month-old mice). Bar graphs correspond to the dark (active) period. (**N**) Kaplan-Meier survival curves for *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 36 to 38 mice, including males and females). Dots in all panels represent individual sample data. Data are presented as means ± SEM. Statistical analysis was by one-way analysis of variance (ANOVA) with post hoc Tukey's correction [(A), *mtCo1* and *mtNd1*; (C), (D), and (G), TNF-α]; Kruskal-Wallis H test with post hoc Dunn's correction [(A), *Tfam*; (B) and (G), IL-6]; unpaired Student's t test [(I), MCV; (J), (K), and (M)]; or unpaired Welch's t test [(L) and (I), hemoglobin]. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001. Survival curve data were analyzed by log-rank (Mantel-Cox test) [(F) and (N)]. For the ECTV experiment, data correspond to one of two representative experiments.

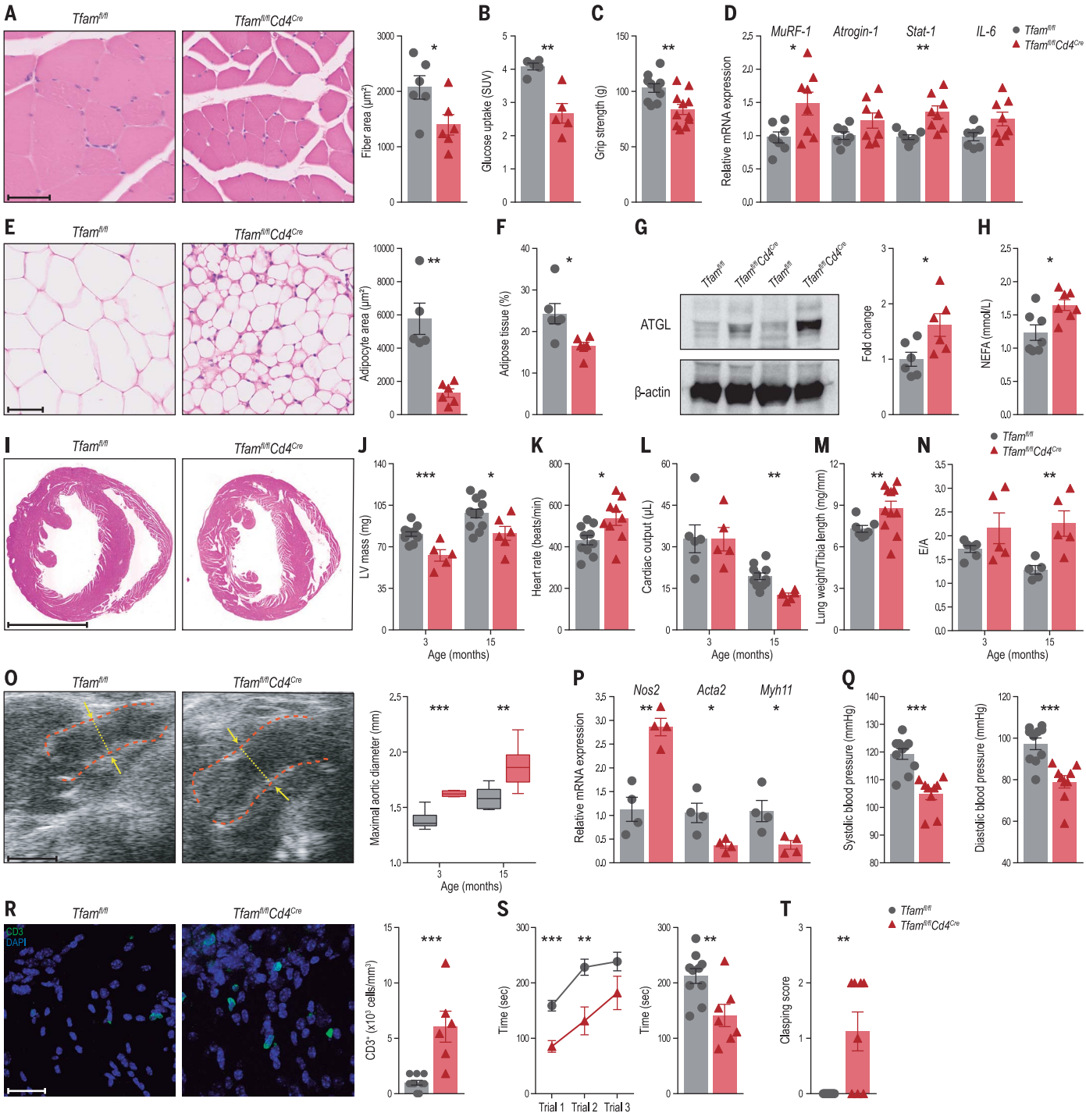


Fig. 2. *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice develop age-associated multimorbidity. (**A**) Representative hematoxylin and eosin (H&E)–stained sections of the gastrocnemius muscle (left; scale bar, 50 μm) and quantification of myofiber cross-sectional area (right). At least 10 measurements were performed per animal. The graph shows mean fiber area for n = 6 animals per group (7-month-old mice). (**B**) In vivo positron emission tomography and CT analysis of skeletal muscle glucose uptake. Data are means ± SEM of ^{18}F-FDG activity (n = 5 4-month-old mice). SUV, standard uptake value. (**C**) Forelimb grip strength analysis (n = 10 to 11 7-month-old mice). (**D**) Relative mRNA levels of genes related to muscle proteolysis [*Murf1* and Atrogin-1 (*Fbxo32*)] and inflammation (*Stat1* and *Il6*) (n = 7 to 8 7-month-old mice). (**E**) Representative H&E-stained sections of gWAT from *Tfam*$^{fl/fl}$ *Cd4*^{Cre} and *Tfam*$^{fl/fl}$ mice (left; scale bar, 50 μm). The graph (right) shows mean estimated adipocyte surface area. Ten measurements were performed per animal (n = 5 to 6 animals, 7-month-old mice). (**F**) Percentage of adipose tissue determined by quantitative magnetic resonance imaging in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} and *Tfam*$^{fl/fl}$ mice (n = 6 7-month-old mice). (**G**) Immunoblot (left) and densitometry analysis (right) of adipose triglyceride lipase (ATGL) protein expression in gWAT isolated from *Tfam*$^{fl/fl}$ *CD4*^{Cre} mice and control littermates (n = 6 7-month-old mice). (**H**) Plasma nonesterified fatty acids (NEFA) (n = 7 4-month-old mice). (**I**) Representative H&E-stained heart sections from 15-month-old *Tfam*$^{fl/fl}$ *Cd4*^{Cre} and *Tfam*$^{fl/fl}$ mice. Scale bar, 2.5 mm. (continued on next page)

(**J**) Echocardiography measurements of left ventricular (LV) mass in 3- and 15-month-old *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 5 to 12). (**K**) Heart rate in 3-month-old *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice and control littermates (n = 9 to 10). (**L**) Cardiac output in 3- and 15-month-old *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 5 to 11). (**M**) Lung weight normalized to tibia length in *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 7 to 11 7-month-old mice). (**N**) Mitral flow pattern on echocardiography in 3- and 15-month-old *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 5 to 6). E/A, E wave/A wave ratio. (**O**) Representative ultrasound images depicting maximal ascending aorta diameters in 15-month-old *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (left; scale bar, 1 mm) and quantification of maximal aortic diameter in 3- and 15-month-old *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (right) (n = 6 to 8). Maximal aortic diameter is presented in box-and-whisker plots showing the median, maximal, and minimal values and 75th and 25th percentiles. (**P**) Real-time quantitative polymerase chain reaction (RT-qPCR) analysis of *Nos2*, *Acta2*, and *Myh11* mRNA expression levels in aortic samples from 7-month-old *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 4). (**Q**) Systolic and diastolic blood pressure in 3-month-old *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 9 to 10). (**R**) Representative CD3-stained brain (fornix) sections from 15-month-old *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (left; scale bar, 25 μm) and quantification of CD3 positive cell density (right) (n = 6 to 10). DAPI, 4′,6-diamidino-2-phenylindole. (**S**) Rotarod test performance by *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice, expressed as the mean time spent on the rotating rod in each of three trials (left) and in all trials combined (right) (n = 7 to 10 12-month-old mice). (**T**) Maximum clasping score per 30-s test (n = 8 to 9 12-month-old mice). Dots in all panels represent individual sample data. Data are presented as means ± SEM. Statistical analysis was by unpaired Student's t test [(A) to (C) and (D), *Murf1*, *Fbxo32*, and *Il6*; (F) to (L), (N), and (O), 15 months; (P) and (Q), diastolic pressure; and (S)]; unpaired Welch's t test [(D), *Stat1*; (M), (N), and (O), 3 months]; or nonparametric Mann-Whitney U test [(Q), systolic pressure; (R) and (T)]. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

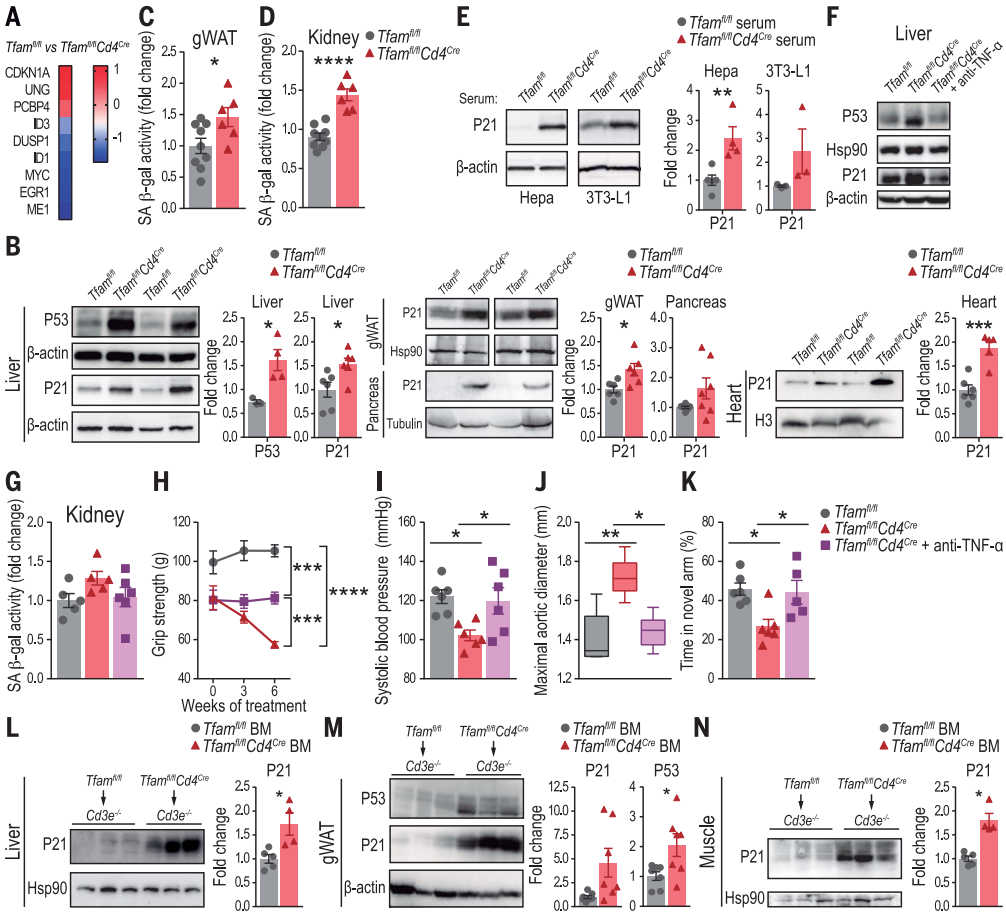


Fig. 3. Inflammaging induces senescence in distal tissues of *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice. (**A**) Heatmap of senescence gene expression changes comparing livers from *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice with those from control littermates. (**B**) Representative immunoblot (left) and quantification (right) of p21 and p53 protein expression in the liver, heart, gWAT, and pancreas from *Tfam*$^{fl/fl}$ *Cd4*^{Cre} and *Tfam*$^{fl/fl}$ mice. Loading controls were β-actin, H3, Hsp90, or α-tubulin (n = 4 to 7 7-month-old mice). (**C**) Quantitative measurements of senescence-associated β-galactosidase (SA β-gal) activity in gWAT lysates by colorimetric assay (n = 6 to 9 9-month-old mice). (**D**) Quantitative measurements of SA β-gal activity in kidney lysates by colorimetric assay (n = 6 to 9 12-month-old mice). (**E**) Immunoblot (left) and densitometry analysis (right) of p21 expression in immortalized mouse hepatocytes (Hepa) and 3T3-L1 cells cultured for 7 days in the presence of *Tfam*$^{fl/fl}$ or *Tfam*$^{fl/fl}$ *Cd4*^{Cre} serum from 7-month-old mice. β-actin was used as a loading control (n = 4 to 6). Blots are representative of three (3T3-L1) or four (hepatocytes) experiments using pooled sera from three animals. (**F**) Representative immunoblot analysis of senescence markers in the liver from control *Tfam*$^{fl/fl}$ and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice with or without treatment with anti–TNF-α (etanercept) (n = 6, 10 weeks of treatment starting from 4 months of age). (**G**) SA β-gal activity measured by a colorimetric assay in kidney lysates (n = 6, 10 weeks of treatment). (**H**) Time course of forelimb strength during anti–TNF-α treatment. (**I**) Systolic blood pressure after 7 weeks of anti–TNF-α treatment (n = 6). (**J**) Maximal ascending aorta diameter in response to anti–TNF-α treatment (n = 6, 8 weeks of treatment). Maximal aortic diameter is presented in box-and-whisker plots showing maximal and minimal values and 75th and 25th percentiles. (**K**) Y-maze analysis in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice treated with anti–TNF-α and corresponding controls (n = 6, 8 weeks of treatment). (**L** to **N**) Immunoblot (left) and densitometry analysis (right) of p21 or p53 expression in liver (L), gWAT (M), and tibialis muscle (N) from *Cd3e*$^{-/-}$ mice 16 weeks after reconstitution with BM from *Tfam*$^{fl/fl}$ *Cd4*^{Cre} or *Tfam*$^{fl/fl}$ mice (n = 4 to 6 per group). Dots in all panels represent individual sample data. Data are presented as means ± SEM. Statistical analysis was by one-way ANOVA with post hoc Tukey's correction [(G), (I), and (K)]; Kruskal-Wallis test with post hoc Dunn's correction (J); two-way ANOVA with post hoc Tukey's correction (H); unpaired Student's t test [(B), (C), (D), (E), (L), and (M)]; or nonparametric Mann-Whitney U test (N). $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

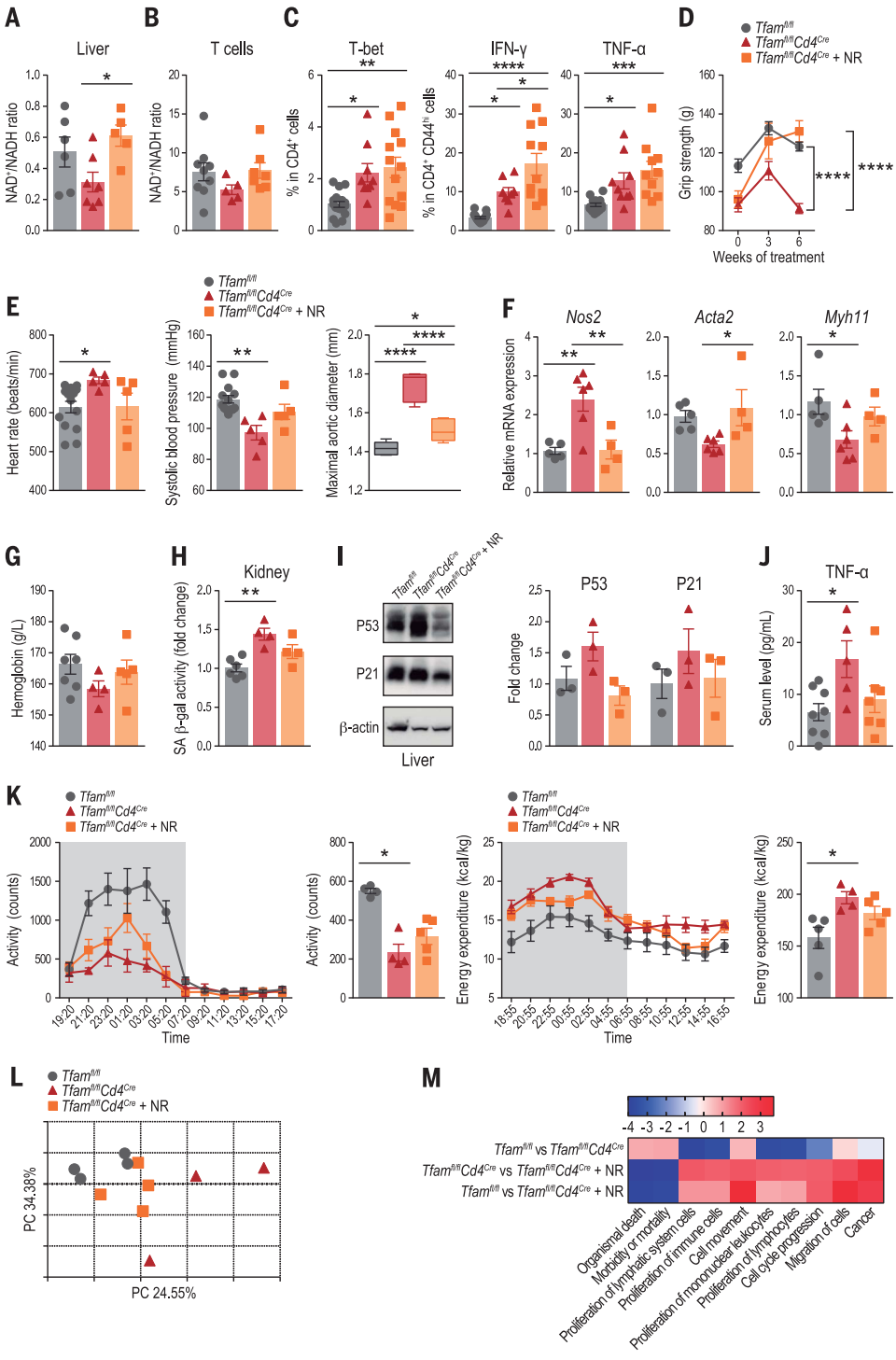


Fig. 4. NR treatment rescues the multimorbidity syndrome. (**A** and **B**) NAD^+/NADH ratio in liver lysates (A) and in isolated $CD4^+$ T cells (B) from untreated *Tfam*$^{fl/fl}$, *Tfam*$^{fl/fl}$ *Cd4*^{Cre}, and NR-treated *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 5 to 9 per group, 10-week treatment starting from 4 months of age). (**C**) Flow cytometry analysis of intracellular T-bet (left), IFN-γ (center), and TNF-α (right) in splenic $CD4^+$ T cells (n = 8 to 17 mice). (**D**) Effect of NR treatment on the evolution of forelimb strength (n = 5 to 11). (**E**) Heart rate (left), systolic blood pressure (center), and ascending aorta (AsAo) maximum diameter (right) obtained from ultrasound images after 8 weeks of NR treatment. Maximal aortic diameter is presented in box-and-whisker plots showing the median and the 25th and 75th percentiles (n = 5 to 14). (**F**) RT-qPCR analysis of *Nos2*, *Acta2*, and *Myh11* mRNA expression in the aorta from untreated *Tfam*$^{fl/fl}$, *Tfam*$^{fl/fl}$ *Cd4*^{Cre}, and NR-treated *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 4 to 6). (**G**) Blood hemoglobin levels (n = 4 to 7). (**H**) SA β-gal activity measured by colorimetric assay in kidney lysates (n = 4 to 6). (**I**) Immunoblot (left) and densitometry analysis (right) of p21 and p53 expression in liver lysates from untreated *Tfam*$^{fl/fl}$, *Tfam*$^{fl/fl}$ *Cd4*^{Cre}, and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice treated with NR for 10 weeks (n = 3 to 4). β-actin was used as a loading control. (**J**) Circulating TNF-α determined by multiplex analysis in serum from untreated *Tfam*$^{fl/fl}$, *Tfam*$^{fl/fl}$ *Cd4*^{Cre}, and NR-treated *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (n = 5 to 8). (**K**) Activity (left) and energy expenditure (right) monitored over a 24-hour cycle in metabolic cages (n = 4 to 5). (**L**) Principal components (PC) analysis of transcriptomics in liver samples from untreated *Tfam*$^{fl/fl}$, *Tfam*$^{fl/fl}$ *Cd4*^{Cre}, and *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice treated with NR for 10 weeks (n = 3 to 4). (**M**) IPA heatmap showing transcriptionally altered cellular pathways. Dots in all panels represent individual sample data. Data are presented as means ± SEM. Statistical analysis was by one-way ANOVA with post hoc Tukey's correction [(A) to (C) and (E), maximal aortic diameter; (G), (H), and (I), p21; (J) and (K), energy expenditure]; Kruskal-Wallis H test with post hoc Dunn's correction [(E), heart rate and systolic blood pressure; (F) and (I), p53; and (K), activity]; or two-way ANOVA with correction post hoc Tukey's correction (D). $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

Diastolic dysfunction was also evident from left ventricular relaxation defects detected in the mitral flow pattern (Fig. 2N). *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice developed an age-dependent aortic dilation (Fig. 2O), which is correlated with increased mRNA expression of inducible nitric oxide synthase (*Nos2*), decreased mRNA expression of the contractile markers smooth muscle actin (*Acta2*) and smooth muscle–specific myosin heavy chain (*Myh11*), and decreased blood pressure (Fig. 2, P and Q). Consistent with these findings, 77% of *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice exhibited aortic regurgitation compared with 16% of controls. Histological analysis of the aortas allowed us to identify aortic dissections associated with inflammatory foci in 50% of *Tfam*$^{fl/fl}$ *Cd4*^{Cre} animals (fig. S2I). Thus, *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice appear to develop cardiac atrophy, overt heart failure, and severe cardiovascular alterations, all of which can precipitate death. *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice also showed signs of neurological disability, including an influx of T cells in the fornix region of the brain and defects in motor coordination in both the rotarod and the tail suspension tests (Fig. 2, R to T). Together, these data support a role for T cells beyond host defense (*18*, *19*) and indicate that the metabolic fitness of T cells is critical for organismal homeostasis.

To verify whether this multimorbidity phenotype was caused by a mitochondrial failure in T cells, we used an alternative approach to delete *Tfam* in T cells. We generated *Tfam*$^{fl/fl}$ *Lck*^{Cre} mice, in which Cre recombinase is expressed under the control of the lymphocyte protein tyrosine kinase (*Lck*) promoter. *Tfam*$^{fl/fl}$ *Lck*^{Cre} mice presented comparable metabolic and functional alterations in T cells and similar premature age-associated multimorbidity to those observed in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (fig. S3). To identify the molecular mechanism by which T cell metabolic failure drives organismal frailty and multimorbidity, we performed liver transcriptomics. *Tfam*$^{fl/fl}$ *Cd4*^{Cre} livers showed significant up-regulation of genes associated with senescence (fig. S4A). One of the most up-regulated of these transcripts was *Cdkn1a*, which encodes the cyclin inhibitor p21 (*Q* value = 0.003) (Fig. 3A). *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice showed elevated protein levels of senescence markers p21$^{Waf/Cip1}$ and p53 in the liver. p21 was also elevated in the heart, gWAT, and pancreas (Fig. 3B). Additionally, the activity of senescence-associated β-galactosidase was higher in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} gWATs and kidneys than in those of controls (Fig. 3, C and D, and fig. S4, B and C). Analysis of *Tfam*$^{fl/fl}$ *Lck*^{Cre} mice confirmed senescence induction in various tissues (fig. S3M).

Because the in vitro incubation of cancer cells with the type 1 cytokines has been shown to induce senescence (*20*), we hypothesized that the type 1 cytokines present in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice drive systemic senescence. Incubation of mouse cells with *Tfam*$^{fl/fl}$ *Cd4*^{Cre} serum or TNF-α was sufficient to increase p21$^{Waf/Cip1}$ levels in hepatocytes and preadipocytes, which supports the argument that inflammatory mediators induce senescence and premature aging (Fig. 3E and fig. S4D). To dissect the contribution of TNF-α to the multimorbidity phenotype, we treated *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice with the TNF-α inhibitor etanercept. Blocking TNF-α prevented systemic senescence (Fig. 3, F and G) and muscle, cardiovascular, and cognitive alterations (Fig. 3, H to K). Bone marrow (BM) transplantation from *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice into T cell–deficient (*Cd3e*$^{-/-}$) mice (fig. S5, A and B) recapitulated the type 1 cytokine storm observed in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (fig. S5, C to E) and signs of senescence in multiple tissues (Fig. 3, L to N). *Cd3e*$^{-/-}$ mice reconstituted with *Tfam*$^{fl/fl}$ *Cd4*^{Cre} BM, but not with control BM, developed a prematurely aged appearance after 16 weeks (fig. S5F) and reproduced the multimorbidity phenotype (fig. S5, G to J). Thus, T cells with mitochondrial dysfunction are able to induce systemic senescence in peripheral tissues by means of a type 1 cytokine storm.

Notably, the senescence observed in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice was accompanied by a low ratio of oxidized nicotinamide adenine dinucleotide (NAD^+) to reduced nicotinamide adenine dinucleotide (NADH) in peripheral tissues (Fig. 4A and fig. S6A). This was consistent with the reported decline in NAD^+ levels during aging (*21*). NAD^+ is a metabolic cofactor with a critical role in mitochondrial function, and restoring NAD^+ levels confers protection against age-associated diseases (*21*–*25*). Recent data suggest that supplementation with the NAD^+ precursor nicotinamide riboside (NR) reduces inflammaging in the elderly (*26*). Accordingly, we observed that NR decreased the serum levels of TNF-α in aged mice (fig. S6B). To test whether NAD^+-boosting strategies could prevent tissue damage caused by inflammaging, we treated *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice with NR for 10 weeks to restore the NAD^+/NADH ratio (Fig. 4, A and B). Although NR neither prevented weight loss nor reversed the T_H1 cell phenotype (Fig. 4C), it successfully prevented skeletal muscle wasting and anemia. Furthermore, most of the cardiovascular alterations found in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice were also prevented (Fig. 4, D to G). NR treatment also increased physical activity levels of *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice and prevented tissue senescence and inflammation (Fig. 4, H to K, and fig. S6C). Finally, a principal components analysis of liver transcriptomics classified *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice treated with NR closer to controls than to untreated *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice (Fig. 4L). Ingenuity pathway analysis (IPA) linked NR treatment to pathways associated with inflammation, morbidity, and mortality (Fig. 4M and table S1). Thus, NR can prevent transcriptional changes related to aging in *Tfam*$^{fl/fl}$ *Cd4*^{Cre} mice.

With the continuous extension of life expectancy, there is an urgent need to understand the common molecular pathways by which aging results in a progressively higher susceptibility to diseases (*27*). Our results indicate that metabolic changes in the immune system promote age-related deterioration in other tissues, which leads to multimorbidity and premature death. Dysregulated T cell metabolism triggers a type 1 cytokine storm that induces senescence in several tissues. Furthermore, we have used this model of multimorbidity and premature aging to test for drugs that can delay aging signs. We found that NAD^+ precursors can prevent the tissue damage associated with sustained inflammation, which supports their potential for preventing age-associated multimorbidity. Further investigation is needed to understand whether targeting senescence or boosting NAD^+ levels could have beneficial effects beyond age-associated diseases in patients with cachexia or cytokine-release syndrome. Our results place immunometabolism at the crossroads of inflammation, senescence, and aging, thereby highlighting its potential as a therapeutic target for delaying aging and aging-associated diseases.

REFERENCE AND NOTES

1. T. N. Tarasenko *et al*., *Cell Metab.* **25**, 1254–1268.e7 (2017).
2. S. E. Weinberg *et al*., *Nature* **565**, 495–499 (2019).
3. L. A. Sena *et al*., *Immunity* **38**, 225–236 (2013).
4. E. L. Pearce *et al*., *Nature* **460**, 103–107 (2009).
5. E. L. Mills, B. Kelly, L. A. J. O'Neill, *Nat. Immunol.* **18**, 488–498 (2017).
6. J. P. Rhoads, A. S. Major, J. C. Rathmell, *Nat. Rev. Rheumatol.* **13**, 313–320 (2017).
7. C. H. Chang, E. L. Pearce, *Nat. Immunol.* **17**, 364–368 (2016).
8. L. Li *et al*., *Cell Metab.* **29**, 103–123.e5 (2019).
9. G. S. Hotamisligil, *Immunity* **47**, 406–420 (2017).
10. A. Wang, H. H. Luan, R. Medzhitov, *Science* **363**, eaar3932 (2019).
11. N. Ron-Harel *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 13347–13352 (2018).
12. M. I. Ekstrand *et al*., *Hum. Mol. Genet.* **13**, 935–944 (2004).
13. F. Baixauli *et al*., *Cell Metab.* **22**, 485–498 (2015).
14. L. Ferrucci, E. Fabbri, *Nat. Rev. Cardiol.* **15**, 505–522 (2018).
15. C. Franceschi, P. Garagnani, P. Parini, C. Giuliani, A. Santoro, *Nat. Rev. Endocrinol.* **14**, 576–590 (2018).
16. D. Furman *et al*., *Nat. Med.* **25**, 1822–1832 (2019).
17. N. M. Chapman *et al*., *Nat. Commun.* **9**, 2095 (2018).
18. S. He *et al*., *Nature* **566**, 115–119 (2019).
19. L. C. Rankin, D. Artis, *Cell* **173**, 554–567 (2018).
20. H. Braumüller *et al*., *Nature* **494**, 361–365 (2013).
21. A. P. Gomes *et al*., *Cell* **155**, 1624–1638 (2013).
22. E. Verdin, *Science* **350**, 1208–1213 (2015).
23. H. Zhang *et al*., *Science* **352**, 1436–1443 (2016).
24. E. Katsyuba *et al*., *Nature* **563**, 354–359 (2018).
25. S. J. Mitchell *et al*., *Cell Metab.* **27**, 667–676.e4 (2018).
26. Y. S. Elhassan *et al*., *Cell Rep.* **28**, 1717–1728.e6 (2019).
27. I. Bellantuono, *Nature* **554**, 293–295 (2018).

ACKNOWLEDGMENTS

We thank N. G. Larsson for *Tfam*$^{fl/fl}$ mice; R. Tejedor and J. L. de Pablos for technical support; and C. López-Otín, M. Serrano, and B. Alarcón for scientific discussions. We also thank A. de Francisco, Y. Sierra, and M. de la Jara Felipe for their excellent work with animal preparation and imaging protocols and D. Calle for her help with data processing. **Funding:** This study was supported by the Fondo de Investigación Sanitaria del Instituto de Salud Carlos III (PI16/02188 and PI19/00855; and PI16/02110 to B.I.), the European Regional Development Fund (ERDF), and the European Commission through H2020-EU.1.1 and European Research Council grant ERC-2016-StG 715322-EndoMitTalk. This work was partially supported by Comunidad de Madrid (S2017/BMD-3867 RENIM-CM). M.M. is supported by the Miguel Servet Program (CPII 19/00014). G.S.-H. is supported by FPI-UAM, J.O. (FJCI-2017-33855) and E.G.-R. (IJC2018-036850) by Juan de la Cierva, and E.C. by Atracción de Talento Investigador 2017-T2/BMD-5766 (Comunidad de Madrid and UAM). B.I. was supported by ERC research grant ERC-2018-CoG 819775-MATRIX. **Author contributions:** G.D.-M., G.S.-H., J.F.A., J.O., E.C., E.G.-R., M.N.N., L.C., F.B., E.M.B., and B.H. performed experimental work and analyzed the data. B.I., A.Alf., M.D., A.R.G., P.F.-M., and A.Alc. provided technical and scientific support. F.B. and M.M. designed the research. G.D.-M. and M.M. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are presented in the main text or the supplementary materials. The *Tfam*$^{fl/fl}$ mice used in this work were obtained under a material transfer agreement with Max Planck Institute for Biology of Aging.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1371/suppl/DC1
Materials and Methods
Figs. S1 to S6
Table S1
References (*28*–*30*)

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

21 February 2019; resubmitted 16 January 2020
Accepted 21 April 2020
Published online 21 May 2020
10.1126/science.aax0860

CHEMICAL ECOLOGY

Information arms race explains plant-herbivore chemical communication in ecological communities

Pengjuan Zu[1]*, Karina Boege[2], Ek del-Val[3], Meredith C. Schuman[4,5], Philip C. Stevenson[6,7], Alejandro Zaldivar-Riverón[8], Serguei Saavedra[1]

Plants emit an extraordinary diversity of chemicals that provide information about their identity and mediate their interactions with insects. However, most studies of this have focused on a few model species in controlled environments, limiting our capacity to understand plant-insect chemical communication in ecological communities. Here, by integrating information theory with ecological and evolutionary theories, we show that a stable information structure of plant volatile organic compounds (VOCs) can emerge from a conflicting information process between plants and herbivores. We corroborate this information "arms race" theory with field data recording plant-VOC associations and plant-herbivore interactions in a tropical dry forest. We reveal that plant VOC redundancy and herbivore specialization can be explained by a conflicting information transfer. Information-based communication approaches can increase our understanding of species interactions across trophic levels.

Chemical information is an ancient and ubiquitous channel to mediate species interactions (*1*) (e.g., attracting or repelling individuals) and is regarded as one of the main forces shaping plant-herbivore interaction networks (*2*, *3*). For example, insects have a large number of olfactory receptors with high sensitivity for chemical signals (*4*, *5*), which play essential roles in their fundamental activities (*6*), including foraging (*7*) and oviposition (*8*). Chemically mediated species interactions are based on information transfer, which we define simply as communication regardless of the benefit to the emitter and receiver [following previous work (*9*)]. Understanding the chemical communication between plants and herbivores has been an active field of research (*10*–*12*), and most work has focused on how a specific plant (or genus) defends against herbivores directly by using chemical repellents (*13*) or by producing stronger chemical defenses (*14*–*16*), or indirectly by emitting signals to attract herbivores' enemies (e.g., predators or parasitoids) (*17*–*22*). So far, only a few descriptive studies (*23*, *24*) have begun to investigate this chemical communication at the community level. More critically, little is known about how information transfer shapes species interactions. For example, it remains unclear how plants code chemical information to deal with a diversity of potential herbivores and how herbivores decode such olfactory signals to distinguish among plants and identify potential hosts. Even less is known about whether there is any general information structure of plant-herbivore chemical communication and how such a structure can be maintained under an ongoing chemical "arms race" between plants and herbivores.

Information theory (*25*), which provides a quantitative and scalable way to measure information transfer, has already brought key insights about the structure and emergence of human language (*26*), and it can be extended to increase our understanding about the "chemical language" in ecological communities (*12*, *27*–*29*). However, such attempts have only been used to study chemical communication of a single plant species with its interacting insects (*30*). Here, we used information theory to study plant volatile organic compounds (VOCs) as a communication channel forming plant-insect interaction networks (*31*). From an information perspective, the relationship between sender and receiver (or speaker and hearer) determines the nature of the signal transmission and the evolution of the information structure shaping the communication pattern between individuals. That is, individuals can either provide clear information that can be decoded easily or spurious information that can be difficult to decode.

Moreover, a conflicting or harmonious communication process between sender and receiver can be modeled by the combinations of maximizing or minimizing uncertainty (or mutual information) of the two parties, respectively (*26*). Specifically, for a given communication system that is formalized by a matrix describing how a vector of signals (S) is associated with a vector of objects (O), the clearness (or the spuriousness) of the communication can be described by two important information measurements (*25*). First, mutual information [$I(O,S)$] describes the amount of information that one can obtain about objects by knowing the signals. Second, conditional entropy [$H(O|S)$] describes the uncertainty of correctly identifying an object given that a specific signal has been observed [for a detailed mathematical account of these expressions, see the supplementary materials (*32*)]. Both mutual information and conditional entropy represent the efficiency of the coding (and decoding) strategies and can be related to the fitness of senders and receivers (*26*).

Tropical dry forests host a large number of interacting and coevolving plant and butterfly species and are regarded as biodiversity hotspots (*33*). Our fieldwork was conducted in a tropical dry forest of the Chamela-Cuixmala Biosphere Reserve (19°22' to 19°39' N; 104°56' to 105°10' W) in Jalisco, Mexico. During the rainy season of 2018, we searched comprehensively for lepidopteran larvae on leaves of target plants in our transect plots and reared the larvae in the laboratory with leaves collected from their host plant species to confirm their trophic interaction and to identify the herbivore species. The supplementary materials (*32*) provide a detailed account of plants and insects, sampling, and identification procedures. We constructed a qualitative (i.e., presence or absence) herbivore (animal)–plant (AP) interaction matrix comprising 28 lepidopteran herbivore species and 20 plant species. The AP matrix was rather sparse (see Fig. 1A and table S1), with a median of one plant species per herbivore (range, one to nine).

Additionally, we sampled the headspace around leaves to retrieve VOCs from each of these plant species. We were able to match 93 analytes from headspace samples to the NIST17 VOC library, of which 56 were likely biogenic (from plants or from plants and microbes), and 31 of these biogenic VOCs could be identified by one or more abundant ions in an untargeted analysis. We used single representative ions from these 31 VOCs for quantification. The supplementary materials (*32*) provide a detailed account of VOC sampling, identification, and quantification procedures. Figure 1B (and table S2) shows that, contrary to the AP matrix, the plant-VOC (PV) association matrix is dense and

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Instituto de Ecología, Universidad Nacional Autónoma de México, 04510 Ciudad de México, México. [3]Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Unidad Morelia Universidad Nacional Autónoma de México, 58190 Morelia, Michoacán, México. [4]Department of Molecular Ecology, Max Planck Institute for Chemical Ecology, DE-07745, Germany. [5]Departments of Chemistry and Geography, University of Zurich, CH-8057 Zurich, Switzerland. [6]Royal Botanic Gardens, Kew Richmond, Surrey TW9 3AB, UK. [7]Natural Resources Institute, University of Greenwich, Chatham, Kent ME4 4TB, UK. [8]Instituto de Biología, Universidad Nacional Autónoma de México, 04510 Ciudad de México, México.
*Corresponding author. Email: pengjuan.zu@gmail.com

erratic, with a median of 26 VOCs per plant (range, 13 to 30).

The *PV* matrix provides information about *PV* associations, whereas the *AV* matrix (mathematically defined by the matrix multiplication $AP \times PV$) provides information about herbivore-VOC associations. These associations describe the patterns of plants emitting information and herbivores receiving it and thus represent plant coding and herbivore decoding strategies. The *AV* matrix (Fig. 1C) is dense and weighted, indicating that individual VOCs are associated with more than one herbivore. Note that to generate the *AV* matrix, we assumed that if a herbivore can feed on a plant, it must be able to cope with spatial and temporal variation in VOCs that are used to distinguish suitable hosts despite the effects of these VOCs on the herbivore (attract, deter, or neutral). Altering either the *PV* or *AP* matrix can change the *AV* matrix.

We built the conceptual framework for our study based on a conflicting optimization process between plants and herbivores (see Fig. 2). Specifically, we hypothesized that plants aim to decrease the decoding efficiency of herbivores by changing *PV* associations, whereas herbivores aim to increase this efficiency by changing *AP* interactions. Ecologically, the decoding strategy can be linked to the fitness of plants and herbivores. A high decoding efficiency can increase the attack rates and decrease the fitness of plants. By contrast, low efficiency can increase the searching time and decrease the fitness of herbivores (*12*). Formally, we define the fitness relationships as $F_P = H(A|V) \in [0,1]$ and $F_A = 1 - H(V|A) \in [0,1]$, respectively. That is, plant fitness is proportional to the conditional entropy (uncertainty) between VOCs and herbivores, whereas herbivore fitness is negatively proportional to the conditional entropy between herbivores and VOCs. The supplementary materials (*32*) provide details on the calculations.

To track the evolutionary trajectory of the communication system between plants and

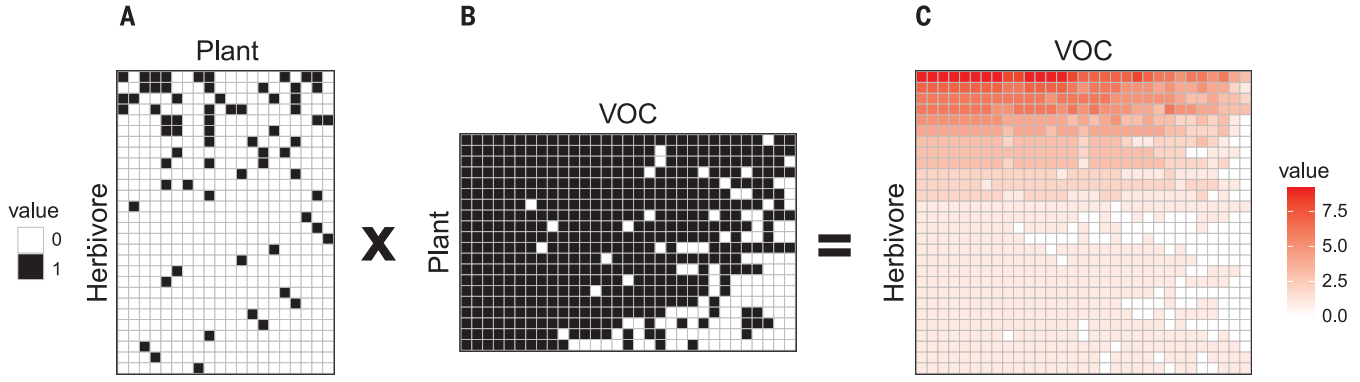


Fig. 1. Field data on *AP* interactions, *PV* associations, and *AV* associations. (**A**) Each row and column represents one of the 28 insect herbivores and 20 plants sampled, respectively. Black and white squares show that a given herbivore was or was not observed feeding on a given plant, respectively. We refer to this as the *AP* matrix and we used it to investigate the plant-herbivore interaction network. (**B**) Each row and column represents one of the 20 plants and 31 VOCs sampled, respectively. Black and white squares show that a given VOC was present or absent in a given plant, respectively. We refer to this as the *PV* matrix and we used it to investigate the coding strategy of plants. (**C**) *AV* matrix produced by the multiplication of the previous two matrices ($AP \times PV$). Each row and column represents one of the 28 herbivores and 31 VOCs associated, respectively. The darker the squares, the greater the frequency with which a given herbivore is associated with a given VOC. We refer to this as the *AV* matrix and we used it to investigate the decoding strategy of herbivores. [See the supplementary materials (*32*) and tables S1 and S2 for more information about these matrices.]

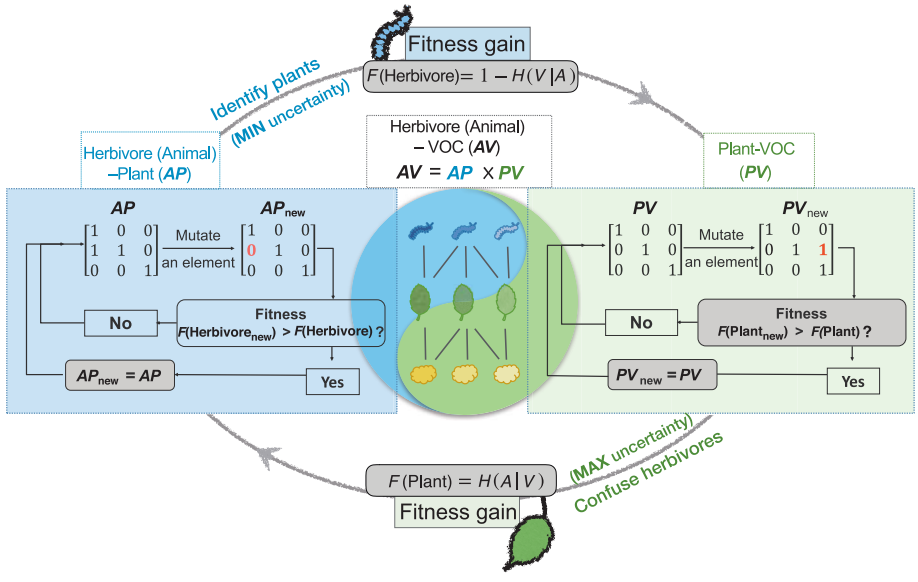


Fig. 2. Conceptual diagram of the proposed conflicting information process (information arms race) between plants and insect herbivores. The coding strategy by plants is characterized by the *PV* association matrix (green right box). The *AP* interaction matrix (blue left box) is used to infer how plants' codes are decoded by herbivores. These matrices undergo mutations by changing any of their elements (e.g., red numbers); zeros and ones correspond to absence and presence, respectively. These mutations will fixate only if the fitness relationship of the corresponding species increases. The fitness relationship of species is defined by herbivores' decoding efficiency based in the communication system, i.e., the *AV* matrix that results from the product of *AP* and *PV* matrices. Specifically, we defined plant and herbivore fitness relationships by the expressions $F_P = H(A|V)$ and $F_A = 1 - H(V|A)$, where $H(\cdot)$ corresponds to the conditional entropy. To increase this fitness relationship, plants and herbivores go through an alternating optimization process, mimicking their arms race. [For more details about this process, see the main text and the supplementary materials (*32*).]

herbivores based on our hypothesis, we simulated the optimization process above in the following way. Plants and herbivores can increase their fitness relationships (based on the *AV* matrix) by modifying the interactions in the *PV* matrix and the *AP* matrix, respectively (see Fig. 2).

These modifications come from random mutations of a given number of elements in the *PV* and *AP* matrices. We assumed an equal mutation rate for each link (*PV* link or *AP* link) in the absence of any prior knowledge. Therefore, the number of random mutations that plants and herbivores could have for each round was proportional to the number of all possible links (i.e., the number of elements in the *PV* and *AP* matrices). Mutations are only selected if they increase the corresponding fitness relationships. To mimic a continuous arms race, we only allowed changes by plants and herbivores in an alternating fashion (see Fig. 2). Both the *PV* and *AP* matrices can be initialized from any random configuration with the only restriction being matrix size. To make our simulation as simple as possible, we assumed that plants can potentially emit any VOC and herbivores can potentially eat any plant without any phylogenetic or physiological constraints.

We investigated the capacity of our hypothesized mechanism to explain both the (de)coding strategies and the information structure of chemical communication between plants and herbivores observed in the field. Specifically, we determined the ability of our model to explain the conditional entropies $H(P|V)$, $H(A|P)$, and $H(A|V)$, which characterize the *PV*, *AP*, and *AV* matrices, respectively. Then, we determined its ability to explain the observed fitness relationships F_P and F_A. Note that conditional entropy by definition is the average entropy given each specific signal [i.e., individual VOC in the case of $H(P|V)$ and $H(A|V)$]. Furthermore, we tested how accurately our model could generate the cumulative information structure given a combination of VOCs by calculating the mutual information between plants and VOCs [$I(P,V)$] and between herbivores and VOCs [$I(A,V)$] as a function of the number of VOCs.

Additionally, we compared our proposed optimization mechanism against three alternative optimization mechanisms. The fitness relationships of plants and herbivores depend on their capacity to (i) maximize uncertainty [i.e., $F_P = H(A|V)$ and $F_A = H(V|A)$], (ii) minimize uncertainty [i.e., $F_P = 1 - H(A|V)$ and $F_A = 1 - H(V|A)$], and (iii) minimize and maximize uncertainty [i.e., $F_P = 1 - H(A|V)$ and $F_A = H(V|A)$]. We also ran all our simulations over different initial conditions, ranging from highly specialized to highly generalized matrices. The rationale for using these alternative mechanisms and initializations was to illustrate that potential matches between our theoretical expectations and the observed values are not just an artifact of sample size, metrics, any given optimization process, or the high (low) connectivity observed in the *PV* (*AP*) matrix. The supplementary materials (*32*) provide all the details about these additional analyses.

Because sampling bias is an important concern in most ecological studies (*34*), we tested whether our results would be affected by the sampling effort in reporting species, by an incomplete plant VOC profile, and by sampling the interaction network in a different year or place. We found that all of our results were qualitatively equivalent when using

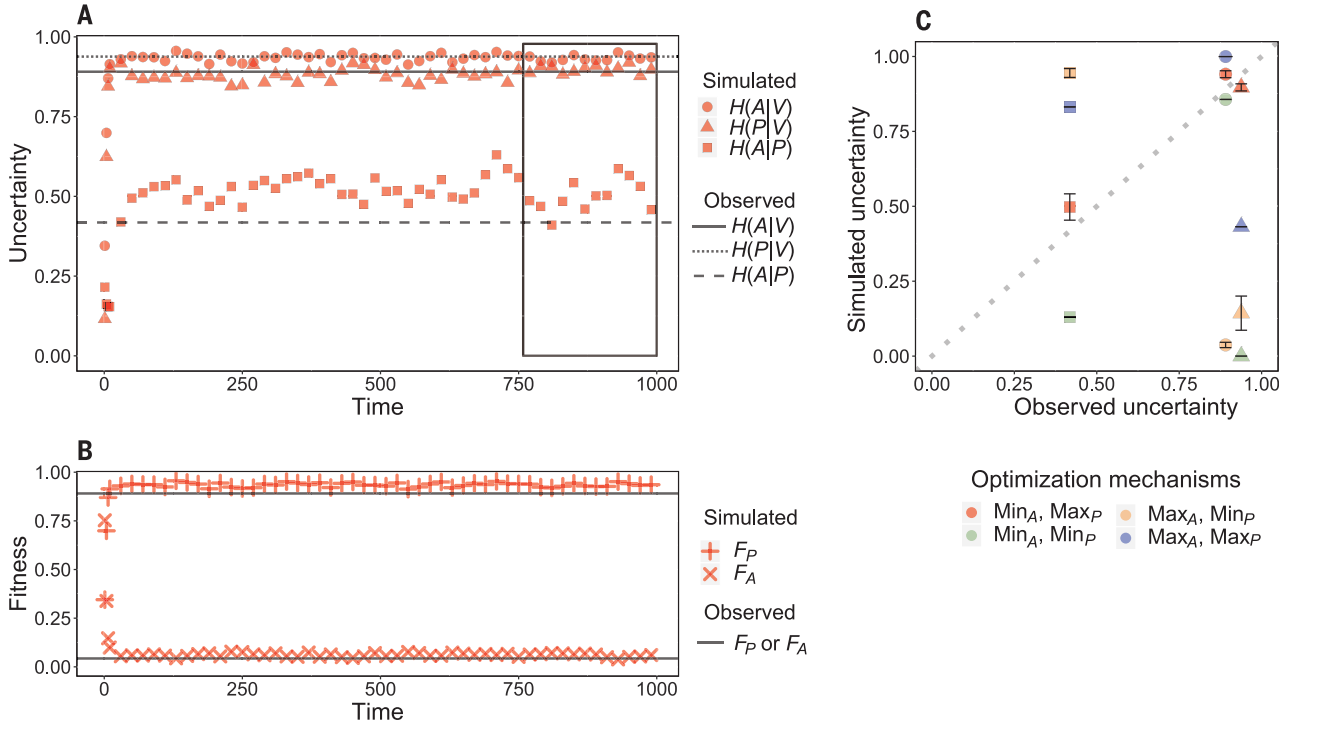


Fig. 3. Simulated results based on the information arms race theory recover the observed information structures in the field. (**A**) Simulation over 1000 time steps derived from our proposed optimization mechanism (information arms race). The simulation is randomly initialized. The circles, triangles, and squares correspond to the conditional entropies $H(A|V)$, $H(P|V)$, and $H(A|P)$, respectively. The solid, dotted, and dashed lines stand for the corresponding conditional entropies from field observations. (**B**) For the same simulation in (A), the + and × symbols correspond to the fitness relationships of plants (F_P) and herbivores (F_A), respectively. The solid lines correspond to field observations. (**C**) Summary of the observed and simulated values (mean ± SD) based on the last 250 steps [the rectangle window in (A) when reaching equilibrium for our proposed mechanism, shown in red; plants aim to maximize ($\max_A$) and herbivores aim to minimize ($\min_P$) uncertainty of information; see Fig. 2] and the corresponding ones for the three alternative mechanisms (in green, yellow, and blue; the alternative combinations of the maximization and minimization).

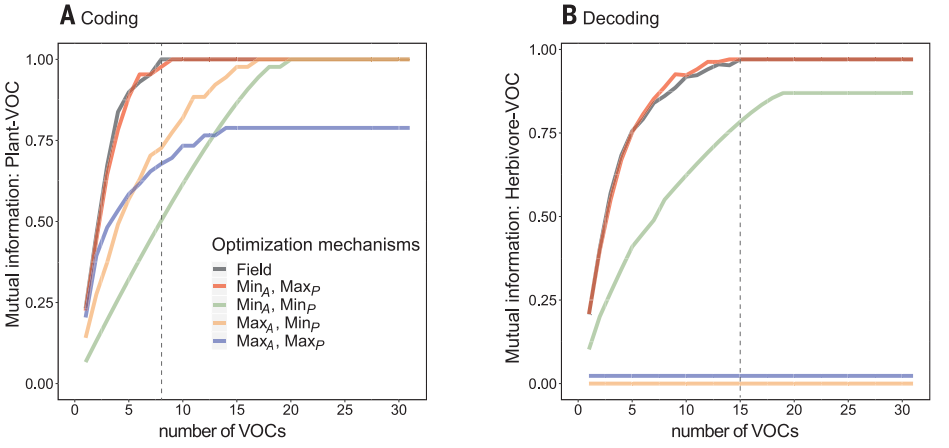


Fig. 4. Emergence of cumulative mutual information. Shown is the information structure characterized by the cumulative mutual information of (**A**) plants and VOCs [i.e., $I(P,V)$] and (**B**) herbivores and VOCs [i.e., $I(A,V)$]. The higher the value, the better the presence or absence of VOCs can identify plants. The cumulative mutual information is calculated as a function of the number of VOCs. The black curve corresponds to the field observation, and the colored curves correspond to the proposed (in red) and alternative optimization mechanisms. Dashed lines correspond to the minimum number of VOCs that gives the best mutual information according to the field data.

subsamples from the whole dataset (illustrating the scalability of our findings), when changing the VOC profile from 31 to 52 by including potential undefined biogenic VOCs, and when using information of plant-herbivore interaction networks from previous years (from 2007 to 2017) (*35*). The supplementary materials (*32*) provide all the details about the robustness of our results to changes in community size, number of VOCs, and identity and number of species interactions.

Figure 3 shows that the proposed conflicting information process is able to explain the coding and decoding patterns observed in the field. In particular, Fig. 3A shows that the conditional entropies $H(P|V)$ (triangles), $H(A|P)$ (squares), and $H(A|V)$ (circles) remain bounded across the simulated time to values close to those observed (see dashed, dotted, and solid lines, respectively). Similarly, Fig. 3B shows that the optimized fitness relationships of plants F_P (+ symbols) and herbivores F_A (× symbols) quickly converge to a stable value close to that observed (solid lines). Figure 3C shows that the alternative optimization mechanisms are not able to explain all the patterns.

Additionally, Fig. 4 shows that of our four tested models, only our proposed conflicting optimization process can recover the information structure observed in the field. Recall that we measured the cumulative information structure by the mutual information [$I(P,V)$ and $I(A,V)$], where values closer to one imply that the presence or absence of a VOC (or a combination of VOCs) gives more information or has higher probability to tell all the species apart. In fact, the field data revealed that as few as eight VOCs can tell all the plants apart and 15 VOCs can give 97% information about herbivores (i.e., have a 97% probability of correctly telling all the herbivores apart). This structure was closely recovered by our proposed optimization mechanism (nine VOCs yield all the information of plants and 14 VOCs yield 97% of information of herbivores; Fig. 4, red lines). By contrast, the three alternative optimization mechanisms yield very different information structures, where either a large number of VOCs would be needed to uniquely identify species or the identification from VOCs would be less likely. For example, in the case where both parties aim to maximize uncertainty (i.e., confuse their opponents as in a mutually competitive relationship; Fig. 4, blue lines), even by using the combination of all VOCs, it would be impossible to tell all the species apart. All these results are robust to changes in community size, number of VOCS, the identity and number of species interactions, and different initial conditions [see the supplementary materials (*32*) and figs. S1 to S6].

Although a few recent studies have begun to demonstrate chemical patterns associated with plant-insect communities (*24*, *31*, *36*), our study proposes a plausible theoretical framework that explains and recovers patterns of information transfer between plants and herbivores from empirical data. Our work is hypothesis driven and suggests that an information arms race between plants and herbivores can drive plants to produce VOCs that are commonly shared by other species, increasing the difficulty for herbivores to identify suitable plants and potentially pressuring herbivores to specialize on a few plants (as the *AP* matrix shows in Fig. 1). Indeed, previous studies have shown a substantial overlap among the VOCs emitted by different plant species (*31*, *37*, *38*). However, the functional roles of a large number of plant VOCs have been unexplained and overlooked because most mechanistic studies have focused on specific plant species and on the ecological functions of only a few VOCs (*12*). Here, our information arms race theory suggests that these seemingly redundant VOCs play an important role in confusing herbivores at the community level. Focusing on herbivore-plant interactions, herbivores are commonly regarded as specialists (*6*), which was also supported by our theory. In fact, it is estimated that <10% of all herbivores feed on plants that span more than three families (*39*).

Our findings further suggest that a conflicting information process drives the rapid accumulation of information by adding VOCs (Fig. 4), contrary to other optimization processes. This provides additional evidence for an arms race process explaining the large diversity of VOCs in nature, where herbivores have the evolutionary potential to quickly tell all plant species apart by making use of the few most informative VOCs, and plants can in turn respond to this potential by adding more VOCs to their profile. Under the same process, herbivores themselves can also be identified using a set of informative VOCs (Fig. 4B). This raises the question of how these VOC profiles result in evolutionary trade-offs regulating the attraction of herbivores and their predators and parasitoids (*17*, *18*). Overall, our study suggests that information transfer processes are key drivers of the formation and maintenance of species interactions across trophic levels.

REFERENCES AND NOTES

1. J. W. Bradbury, S. L. Vehrencamp, *Principles of Animal Communication* (Sinauer, 1998).
2. P. R. Ehrlich, P. H. Raven, *Evolution* **18**, 586–608 (1964).
3. L. A. Richards *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10973–10978 (2015).
4. U. B. Kaupp, *Nat. Rev. Neurosci.* **11**, 188–200 (2010).
5. B. S. Hansson, M. C. Stensmyr, *Neuron* **72**, 698–711 (2011).
6. L. M. Schoonhoven, B. Van Loon, J. J. van Loon, M. Dicke, *Insect-Plant Biology* (Oxford Univ. Press, 2005).
7. F. P. Schiestl, *Ecol. Lett.* **13**, 643–656 (2010).
8. J. Renwick, F. Chew, *Annu. Rev. Entomol.* **39**, 377–400 (1994).
9. R. Karban, *Ecol. Lett.* **11**, 727–739 (2008).
10. T. J. Bruce, J. A. Pickett, *Phytochemistry* **72**, 1605–1611 (2011).
11. M. C. Schuman, I. T. Baldwin, *Annu. Rev. Entomol.* **61**, 373–394 (2016).
12. A. Kessler, A. Kalske, *Annu. Rev. Ecol. Evol. Syst.* **49**, 115–138 (2018).
13. S. S. Duffey, M. J. Stout, *Arch. Insect Biochem. Physiol.* **32**, 3–37 (1996).
14. A. M. Helms, C. M. De Moraes, J. F. Tooker, M. C. Mescher, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 199–204 (2013).
15. N. Bittner, J. Hundacker, A. Achotegui-Castells, O. Anderbrant, M. Hilker, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24668–24675 (2019).

16. D. M. Magalhães, I. T. F. A. Da Silva, M. Borges, R. A. Laumann, M. C. Blassioli-Moraes, . *J. Exp. Bot.* **70**, 1891–1901 (2019).
17. T. C. J. Turlings, M. Erb, *Annu. Rev. Entomol.* **63**, 433–452 (2018).
18. C. M. De Moraes, W. Lewis, P. Pare, H. Alborn, J. Tumlinson, *Nature* **393**, 570–573 (1998).
19. E. H. Poelman, M. Dicke, in *Annual Plant Reviews: Insect-Plant Interactions* (Wiley, 2018), vol. 47, pp 309–337.
20. A. Kessler, I. T. Baldwin, *Science* **291**, 2141–2144 (2001).
21. A. Kessler, M. Heil, *Funct. Ecol.* **25**, 348–357 (2011).
22. J. X. Becerra, K. Noge, D. L. Venable, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18062–18066 (2009).
23. A. Kantsa *et al*., *Nat. Ecol. Evol.* **1**, 1502–1510 (2017).
24. I. Filella *et al*., *Sci. Rep.* **3**, 3434 (2013).
25. C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379–423 (1948).
26. R. Ferrer i Cancho, R. V. Solé, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 788–791 (2003).
27. R. Margalef, *Perspectives in Ecological Theory* (Univ. of Chicago Press, 1968).
28. R. E. Ulanowicz, *Comput. Chem.* **25**, 393–399 (2001).
29. M. I. O'Connor *et al*., *Front. Ecol. Evol.* **7**, 219 (2019).
30. L. Doyle, *Entropy* **11**, 431–442 (2009).
31. G. Vivaldo, E. Masi, C. Taiti, G. Caldarelli, S. Mancuso, *Sci. Rep.* **7**, 11050 (2017).
32. See the supplementary materials.
33. L. Miles *et al*., *J. Biogeogr.* **33**, 491–505 (2006).
34. J. M. Chase, T. M. Knight, *Ecol. Lett.* **16** (suppl. 1), 17–26 (2013).
35. K. Boege *et al*., *Biotropica* **51**, 529–537 (2019).
36. A. Kantsa, R. A. Raguso, T. Lekkas, O. I. Kalantzi, T. Petanidou, *J. Ecol.* **107**, 2574–2586 (2019).
37. E. A. Courtois *et al*., *J. Chem. Ecol.* **35**, 1349–1362 (2009).
38. J. T. Knudsen, R. Eriksson, J. Gershenzon, B. Ståhl, *Bot. Rev.* **72**, 1–120 (2006).
39. E. Bernays, M. Graham, *Ecology* **69**, 886–892 (1988).
40. P. Zu, K. Boege, E. del-Val, M. C. Schuman, P. C. Stevenson, A. Zaldivar-Riverón, S. Saavedra, Data for: Information arms race explains plant-herbivore chemical communication in ecological communities, Zenodo (2020); https://doi.org/10.5281/zenodo.3743129.

ACKNOWLEDGMENTS

We thank I. S. Aranda and R. Moore for help with fieldwork; G. Kite for help with gas chromatography–mass spectrometry; R. Pérez-Ishiwara for logistical support; R. Halitschke at the Max Planck Institute for Chemical Ecology for providing silicon tubing for volatile sampling; M. Al-Adwani, S. Cenci, L. Medeiros, and C. Song for numerous constructive discussions about this work; and R. Karban and R. Ulanowicz for valuable comments on the manuscript. **Funding:** This work was funded by Swiss National Science Foundation grant no. P2ZHP3 178087 to P.Z. E.d.-V. and K.B. were supported by PAPIIT-UNAM (IN211916) and SEP-CONACYT (2015-255544) for fieldwork. S.S. was supported by the Mitsui Chair. M.C.S. was supported by the Max Planck Society, the NOMIS Foundation, and the University of Zurich Research Priority Program on Global Change and Biodiversity. **Author contributions:** P.Z. and S.S. designed the study. P.Z. and E.d.-V. collected field data. K.B., E.d.-V., and A.Z.-R. identified herbivores and constructed plant-herbivore networks. P.Z., M.C.S., and P.C.S. performed chemical analyses. P.Z. and S.S. performed theoretical analyses. P.Z. and S.S. wrote the manuscript with contributions by M.C.S. and all other coauthors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All source data and R code needed to reproduce the results are available in the supplementary materials (*32*) and on Zenodo (*40*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1377/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 and S2
References (*41*–*46*)
Database S1
MDAR Reproducibility Checklist

5 December 2019; resubmitted 25 February 2020
Accepted 15 April 2020
10.1126/science.aba2965

METALLOENZYMES

Structural evidence for a dynamic metallocofactor during N_2 reduction by Mo-nitrogenase

Wonchull Kang[1]*, Chi Chung Lee[1]*, Andrew J. Jasniewski[1], Markus W. Ribbe[1,2]†, Yilin Hu[1]†

The enzyme nitrogenase uses a suite of complex metallocofactors to reduce dinitrogen (N_2) to ammonia. Mechanistic details of this reaction remain sparse. We report a 1.83-angstrom crystal structure of the nitrogenase molybdenum-iron (MoFe) protein captured under physiological N_2 turnover conditions. This structure reveals asymmetric displacements of the cofactor belt sulfurs (S2B or S3A and S5A) with distinct dinitrogen species in the two αβ dimers of the protein. The sulfur-displaced sites are distinct in the ability of protein ligands to donate protons to the bound dinitrogen species, as well as the elongation of either the Mo–O5 (carboxyl) or Mo–O7 (hydroxyl) distance that switches the Mo-homocitrate ligation from bidentate to monodentate. These results highlight the dynamic nature of the cofactor during catalysis and provide evidence for participation of all belt-sulfur sites in this process.

Nitrogenase catalyzes a key step in the global nitrogen cycle: the ambient reduction of atmospheric dinitrogen (N_2) to the bioavailable ammonia (NH_3). The molybdenum (Mo) nitrogenase consists of two component proteins (*1*–*3*). One, designated the iron (Fe) protein, is a homodimer that contains a [Fe_4S_4] cluster at the subunit interface and an adenosine triphosphate (ATP)–binding site within each subunit. The other, designated the molybdenum iron (MoFe) protein, is an $\alpha_2\beta_2$ heterotetramer that contains two complex metalloclusters per αβ dimer: a P-cluster ([Fe_8S_7]) at each interface of the α and β subunits, and an M-cluster {or FeMoco; [(*R*-homocitrate)$MoFe_7S_9C$]} within each α subunit (*4*, *5*). Catalysis by the Mo-nitrogenase involves repeated association and dissociation between the two component proteins, which permits ATP-dependent electron transfer from the [Fe_4S_4] cluster of the Fe protein, via the P-cluster, to the M-cluster of the MoFe protein, where substrate reduction occurs (*1*–*3*). The ability of nitrogenase to shuttle multiple electrons to its cofactor renders the enzyme highly versatile in substrate reduction. Other than N_2 and protons, nitrogenase can reduce a variety of alternative substrates, such as acetylene, cyanide, and carbon monoxide (*1*, *6*, *7*).

Considerable efforts have gone into the mechanistic investigation of nitrogenase since its discovery, with a number of studies focused on the substrate and inhibitor interactions of this enzyme (*1*, *3*, *8*–*10*). However, characterization of the substrate- or intermediate-bound states of nitrogenase has proven to be challenging because of the transient nature of these bound species (*11*). To circumvent this problem, freeze-quench spectroscopic techniques, combined with strategic genetic modifications of the active site, have been applied to accumulate and characterize nitrogenase-bound intermediates, leading to the proposal of a reductive elimination mechanism (*3*) and a concurrent refinement of the Thorneley-Lowe model (*8*, *9*) for N_2 reduction by Mo-nitrogenase. Moreover, high-resolution crystal structures of Mo- and V-nitrogenases have been obtained with ligands (e.g., CO or an N or O species) bound to the enzyme (*12*, *13*), pointing to an intriguing mechanism in which a reactive Fe species is generated upon displacement of a specific belt sulfur (S2B) of the cofactor by a μ_2-bridging ligand. An elegant crystallographic pulse-chase study using a Se reporter further demonstrated a catalysis-dependent migration of belt sulfur (*14*), highlighting the dynamic nature of the entire belt region of the cofactor during turnover. Together, these studies offered invaluable insights into the reaction mechanism of nitrogenase while providing added incentive for exploration of strategies to specifically capture N_2 or related intermediate(s) for mechanistic studies.

One strategy worthy of consideration is to limit excess electron supplies that inadvertently drive the N_2 reduction forward, which either reverts the substrate- or intermediate-bound state to the resting state or renders the enzyme in an indiscernible mixed state. This is particularly relevant because nitrogenase proteins are routinely isolated in the presence of excess dithionite, an externally supplied reductant (*1*). Removal of this artificial electron source, coupled with quick isolation of the protein in the absence of oxygen, could facilitate capture of N_2 or intermediate(s) on the protein upon exhaustion of available electrons. To obtain a proof of concept for this strategy, we first prepared the crude extract of an *Azotobacter vinelandii* strain under anaerobic conditions with or without addition of dithionite upon cell disruption. In both cases, the *A. vinelandii* strain was actively expressing a Mo-nitrogenase comprising a histidine-tagged MoFe protein and a nontagged Fe protein before cell disruption (*15*, *16*). Activity analysis of these samples revealed that, unlike the dithionite-treated crude extract, the dithionite-free crude extract was nearly inactive in substrate reduction, but its activity could be fully restored with the addition of dithionite and ATP (fig. S1A). This observation suggests a depletion of electrons (e.g., from ferredoxins) and other metabolites (e.g., ATP) in the crude extract when the metabolic processes that (re)generate these components are interrupted upon cell disruption. Thus, when a nitrogenase-expressing culture that is actively performing N_2 fixation is subjected to cell lysis without additional electron supplies, the nitrogenase remains functional but is potentially arrested in a "dormant" N_2- or intermediate-bound state owing to a withdrawal of electron flow to the M-cluster. Indeed, after a quick one-step purification under N_2 from the dithionite-free crude extract, the histidine-tagged MoFe protein (designated Av1*) was not only active in N_2 reduction but also fully functional upon substitution of the gas atmosphere with C_2H_2 or Ar in an in vitro activity assay where it was combined with the Fe protein (designated Av2), ATP, and dithionite (fig. S1B). Furthermore, gas chromatography–mass spectrometry analysis demonstrated that, unlike Av1 purified with dithionite from a $^{15}N_2$-grown culture (fig. S1C, black), Av1* purified without dithionite from the same $^{15}N_2$-grown culture released $^{15}N_2$ upon acid quenching (fig. S1C, red), suggesting capture of N_2 or related species on Av1*. Release of $^{15}N_2$ was not detected from the reductant-free resting-state Av1 (designated Av1′) upon incubation with $^{15}N_2$ (fig. S1C, brown), pointing to the turnover conditions as the prerequisite for capturing N_2 on Av1 while providing further support for the physiological relevance of Av1* to catalysis.

Subsequent crystallization of Av1* yielded brown crystals that diffracted to a resolution of 1.83 Å (figs. S2 to S4 and tables S1 to S3). Notably, the two P-clusters in the two αβ dimers of Av1* [designated P-cluster$^{(A/B)}$ and P-cluster$^{(C/D)}$, respectively] are no longer present in the all-ferrous resting state (P^N); instead, both clusters adopt the conformation of the previously reported two-electron oxidized state of this cluster (P^{OX}) (*17*, *18*), with the central S (designated S1) losing two Fe–S bonds and becoming μ_4-coordinated to three Fe atoms of the α-subunit cubane and one Fe atom of

[1]Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA 92697-3900, USA. [2]Department of Chemistry, University of California, Irvine, Irvine, CA 92697-2025, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: mribbe@uci.edu (M.W.R.); yilinh@uci.edu (Y.H.)

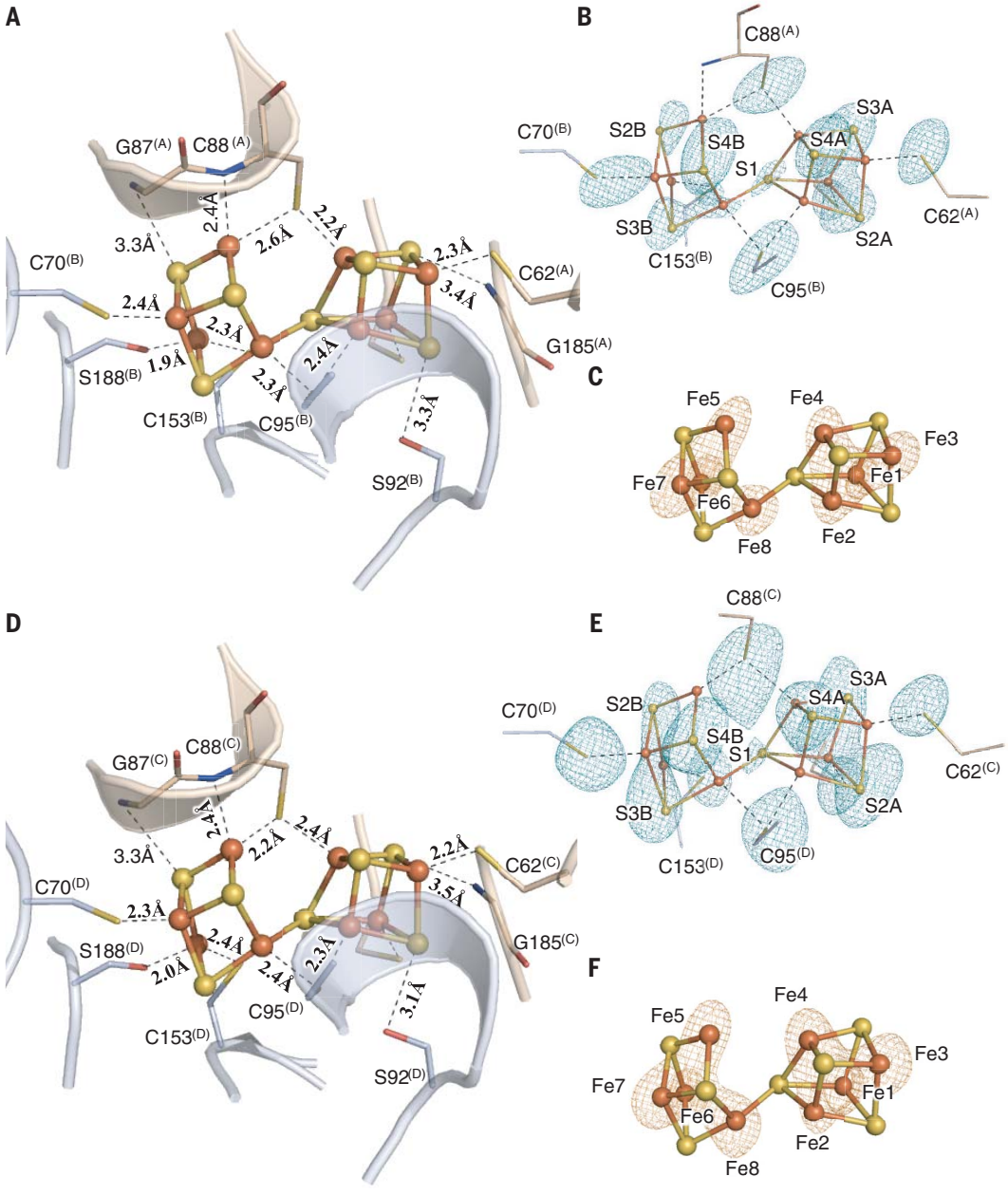


Fig. 1. The oxidized P-clusters in Av1*. Structures of P-clusters at (**A** to **C**) the interface of chains A and B [P-cluster$^{(A/B)}$] and (**D** to **F**) the interface of chains C and D [P-cluster$^{(C/D)}$] of Av1*. Chains A and C are the α subunits, and chains B and D are the β subunits of the two αβ dimers of Av1*. [(A) and (D)] The P-clusters are shown in ball-and-stick presentation, and the key residues interacting with the P-clusters are indicated as sticks. Chains A and C are shown as wheat ribbons, and chains B and D are shown as light-blue ribbons. [(B) and (C)] P-cluster$^{(A/B)}$ and [(E) and (F)] P-cluster$^{(C/D)}$ superimposed with [(B) and (E)] the anomalous density maps calculated at 7100 eV at a resolution of 2.18 Å and contoured at 4.0σ, showing the position of sulfur atoms (mint-blue mesh); and with [(C) and (F)] the anomalous density maps calculated at 7141 eV at a resolution of 2.1 Å and contoured at 15.0σ, showing the position of iron atoms (red mesh). Atoms are colored as follows: Fe, orange; S, yellow; O, red; N, blue. Single-letter abbreviations for the amino acid residues are as follows: C, Cys; G, Gly; H, His; R, Arg; S, Ser.

the β-subunit partial cubane (Fig. 1 and figs. S5 and S6). The structural rearrangement of the P-clusters in Av1*, confirmed by anomalous density data calculated at 7100 and 7141 eV, is accompanied by formation of two Fe–X bonds with an O atom of α-S188 and the backbone N atom of α-C88 (Fig. 1 and figs. S5 and S6). Electron paramagnetic resonance analysis provided further support for this assignment, showing a P^{OX}-specific [proportionality factor (g) = 11.8] signal in the spectrum of Av1* (fig. S1D) (*19*). The P^{OX} signal of Av1* is indistinguishable in intensity from that of an equimolar amount of Av1 treated with excess oxidant, suggesting that all P-clusters of Av1* exist in the P^{OX} state (fig. S1D). The presence of the P-clusters of Av1* in the P^{OX} state not only provides the long-sought-after answer to the physiological relevance of this state, but it also points to a limited electron flow from the P-cluster to the M-cluster owing to the absence of dithionite, leaving the former in an electron-depleted state while permitting the latter to capture N_2 and/or intermediate(s) under limited turnover conditions.

Consistent with this suggestion, in the α subunit (designated chain A) of one αβ dimer of Av1*, the absence of a belt S (S2B) from the M-cluster [designated M-cluster$^{(A)}$] can be clearly and reproducibly visualized in the anomalous difference Fourier map calculated with diffraction data at 7100 eV (Fig. 2, A and B, and fig. S7, A and B); yet, based on analysis of the $2F_o$-F_c and F_o-F_c maps (fig. S8A), there is electron density at this location, which can be modeled as a dinitrogen species that binds in a pseudo $\mu_{1,2}$-bridging mode between Fe2 and Fe6 (Fig. 2C; see figs. S9 and S10 for ligand modeling). The proximal (N_{prox}) and distal (N_{dist}) nitrogen atoms of this species are located at a distance of 1.8 and 2.3 Å, respectively, from Fe2 and Fe6, rendering this dinitrogen species roughly parallel to the M-cluster along the Fe1–C–Mo axis (Fig. 2C and table S3). Additionally, N_{prox} interacts with the Nε group of α-H195$^{(A)}$, a residue implicated in N_2 reduction (*20*), via hydrogen bonding at 2.9 Å (Fig. 2C). Sulfur anomalous density [designated S$^{(A/B)}$] is observed at the same sulfur-binding site (SBS) at the interface of chains A and B as that identified earlier in the S2B-displaced, CO-bound structure of the

M-cluster (*12*). Located ~22 Å away from the S2B site in chain A of Av1* (fig. S11, A and B), the relevance of this sulfur species to the displaced S2B sulfur is yet to be established. Displacement of S2B with the dinitrogen species, like that of S2B with CO, does not have much impact on the structure of the M-cluster, as both ligand-bound clusters show an overall conservation of the Fe–S, Fe–Fe, and Mo–S distances as those in the resting-state cofactor (fig. S12 and table S3) (*5*). However, contrary to its CO-bound and resting-state counterparts (*12*), there is a considerable change in the ligation of Mo by homocitrate in chain A of Av1*. Whereas the Mo–O5 (carboxyl) distance remains largely unchanged, there is a notable elongation of the Mo–O7 (hydroxyl) distance, thereby switching the ligation of Mo by homocitrate from bidentate in the resting-state and CO-bound structures (Mo–O5, 2.2 Å; Mo–O7, 2.2 Å) to monodentate in chain A of Av1* (Mo–O5, 2.3 Å; Mo–O7, 2.7 Å) (Fig. 2C).

Unexpectedly, in the M-cluster [designated M-cluster$^{(C)}$] of the α subunit (designated chain C) of the second αβ dimer, sulfur is clearly and reproducibly visualized at the S2B position but not at the S3A and S5A positions in the anomalous difference Fourier map calculated from diffraction data at 7100 eV (Fig. 2, D and E, and fig. S7, C and D). Like the S2B position in chain A, the electron densities observed at the S3A and S5A positions on the basis of analysis of the $2F_o$-F_c and F_o-F_c maps of chain C (fig. S8B) can each be modeled with a dinitrogen species with N_{prox} and N_{dist} relative to the cluster Fe atoms (Fig. 2F, and figs. S9 and S10). At the S3A site, N_{prox} is located 1.8 and 2.1 Å, respectively, from Fe4 and Fe5; whereas at the S5A site, N_{prox} is located 2.1 and 1.8 Å, respectively, from Fe3 and Fe7 (Fig. 2F and table S3). As such, the dinitrogen species at S3A or S5A appears to form an asymmetric $\mu_{1,1}$ bridge between two Fe centers, with a stronger (1.8 Å) and a weaker (2.1 Å) interaction of N_{prox} with a primary (Fe4 or Fe7) and a secondary (Fe5 or Fe3) Fe center. In both cases, the distal nitrogen atoms are directed away from the M-cluster, rendering the two nitrogen atoms in a somewhat linear alignment with the primary Fe center. Besides interacting with the cluster, the dinitrogen species at the S3A site has potential hydrogen bonding interactions with the amide groups of the protein backbone [α-G356$^{(C)}$ and α-G357$^{(C)}$; N_{dist}–N_{amide} = 2.9–3.4 Å); whereas the dinitrogen species at the S5A site has potential hydrogen bonding interactions with the side chain of α-R96$^{(C)}$ (N_{dist}–N_{R96} = 3.2 Å), a residue implicated in N_2 reduction (*21*), and a water molecule (N_{dist}–O_{H2O} = 3.0 Å) (Fig. 2F). As observed at the interface of chains A and B, sulfur anomalous density [designated S$^{(C/D)}$] appears at the same SBS at the interface of chains C and D, which is ~19 and 25 Å, respectively, away from the S5A and S3A sites (fig. S11, C and D). Additionally, the M-cluster in chain C adopts a conformation that closely resembles the resting-state structure (fig. S13) (*5*). Finally, there is also a clear switch of the ligation of Mo by homocitrate from bidentate to monodentate in chain C. However, the two Mo–O distances in chain C undergo changes opposite to those in chain A, with the Mo–O5 (carboxyl) distance lengthened to 2.7 Å and the Mo–O7 (hydroxyl) distance shortened to 2.0 Å (Fig. 2F).

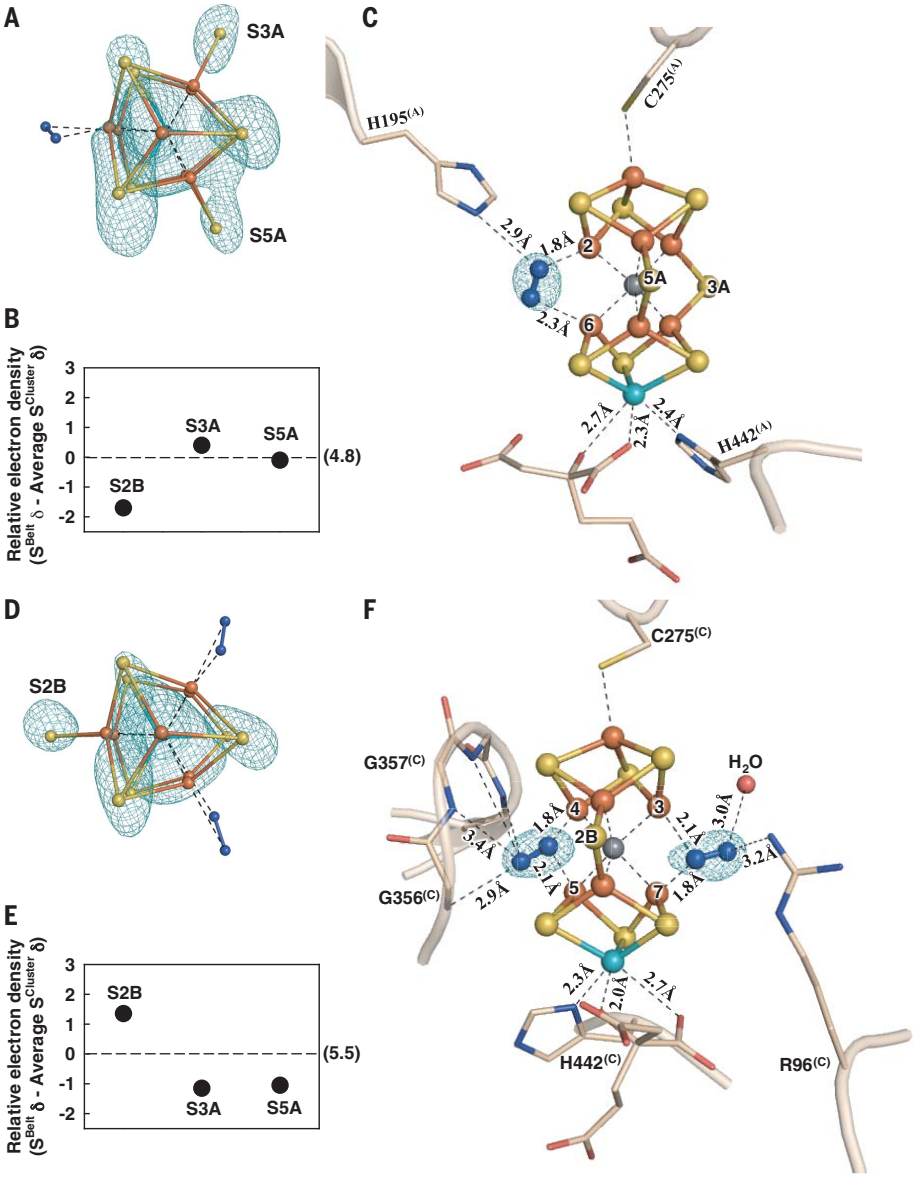


Fig. 2. The nitrogen ligand–bound M-clusters in Av1*. Structures of M-clusters in (**A** to **C**) chain A [M-cluster$^{(A)}$] and (**D** to **F**) chain C [M-cluster$^{(C)}$] refined at a resolution of 1.83 Å. View along the Fe1-C-Mo direction of (A) M-cluster$^{(A)}$ and (D) M-cluster$^{(C)}$ superimposed with the anomalous density maps (mint-blue mesh) calculated at 7100 eV at a resolution of 2.18 Å and contoured at 4.0σ (A) and 5.3σ (D), respectively, showing the displacement of sulfur at the S2B site in M-cluster$^{(A)}$ (A) and at the S3A and S5A sites in M-cluster$^{(C)}$ (D). Electron densities of the belt sulfurs (S^{Belt}) relative to the average density of cluster sulfurs ($S^{Cluster}$) in (B) chain A and (E) chain C, respectively, expressed in sigma values. The average sigma values of the cluster sulfurs in chains A and C are 4.8 and 5.5, respectively [(B) and (E)]. Side view of (C) M-cluster$^{(A)}$ and (F) M-cluster$^{(C)}$ with key residues interacting with the clusters and the bound nitrogen ligands indicated as sticks. M-cluster$^{(A)}$ and M-cluster$^{(C)}$ are superimposed with the F_o-F_c omit map of the nitrogen ligands contoured at 10σ (mint-blue mesh). The peptides and atoms are colored as in Fig. 1.

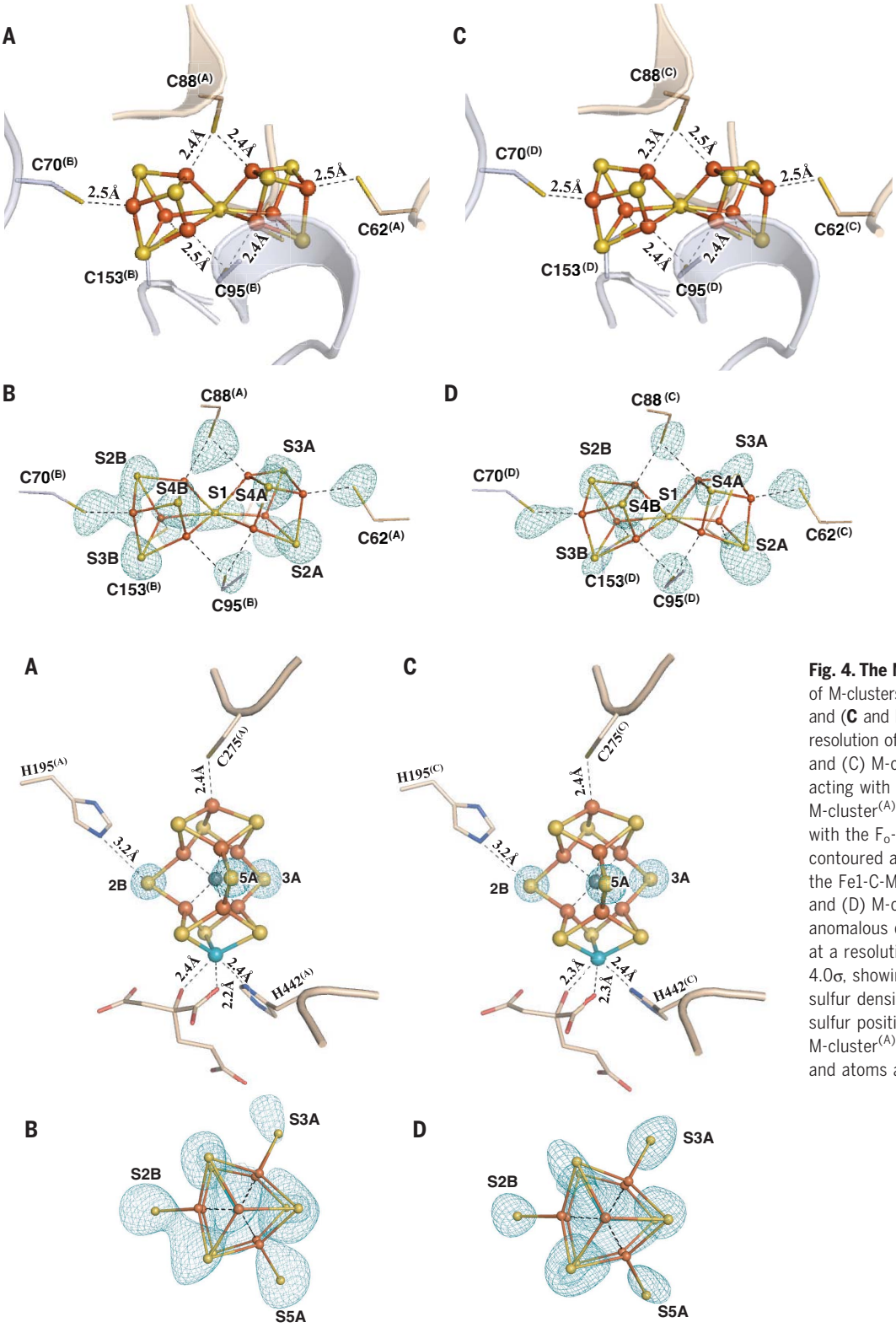


Fig. 3. The reduced P-clusters in Av1*(TOD). Structures of P-clusters at (**A** and **B**) the interface of chains A and B [P-cluster$^{(A/B)}$] and (**C** and **D**) the interface of chains C and D [P-cluster$^{(C/D)}$] of Av1*(TOD). Chains A and C are the α subunits, and chains B and D are the β subunits of the two αβ dimers of Av1*(TOD). [(B) and (D)] P-clusters superimposed with the anomalous density maps calculated at 7100 eV at a resolution of 2.17 Å and contoured at 4.5σ, showing the position of sulfur atoms (mint-blue mesh). The peptides and atoms are colored as in Fig. 1.

Fig. 4. The M-clusters in Av1*(TOD). Structures of M-clusters in (**A** and **B**) chain A [M-cluster$^{(A)}$] and (**C** and **D**) chain C [M-cluster$^{(C)}$] refined at a resolution of 1.73 Å. Side view of (A) M-cluster$^{(A)}$ and (C) M-cluster$^{(C)}$ with key residues interacting with the clusters indicated as sticks. M-cluster$^{(A)}$ and M-cluster$^{(C)}$ are superimposed with the F_o-F_c omit maps of the belt sulfurs contoured at 13σ (mint-blue mesh). View along the Fe1-C-Mo direction of (B) M-cluster$^{(A)}$ and (D) M-cluster$^{(C)}$ superimposed with the anomalous density maps calculated at 7100 eV at a resolution of 2.17 Å and contoured at 4.0σ, showing the presence of the anomalous sulfur density (mint-blue mesh) at all belt-sulfur positions (S2B, S3A, and S5A) in (B) M-cluster$^{(A)}$ and (D) M-cluster$^{(C)}$. The peptides and atoms are colored as in Fig. 1.

To further validate the sulfur-displaced conformation captured in Av1* under a limited electron flux, Av1* was incubated with Av2, ATP, and excess dithionite and reisolated from the incubation mixture. Crystallization of this reisolated Av1 protein [designated Av1*(TOD); TOD, turnover with dithionite] yielded brown crystals that diffracted to a resolution of 1.73 Å (figs. S14 and S15 and tables S1, S3, and S4). Both P-clusters of Av1*(TOD) adopt the conformation of the reduced P^N state (Fig. 3 and figs. S16 and S17); whereas both M-clusters of Av1*(TOD) adopt the resting-state conformation, with all three belt sulfurs in place and homocitrate assuming bidentate ligation to Mo via O5 and O7 (Fig. 4 and figs. S18 and S19). These observations are consistent with the bound dinitrogen species being turned over

and subsequently released from the M-clusters, as well as the inability of the M-clusters to (re)capture N_2 or N_2-derived intermediates in the presence of excess reductant, which inadvertently reverts both P- and M-clusters in Av1*(TOD) to the well-established resting-state conformations. The ability of Av1* to resume turnover further illustrates the physiological relevance of this conformation to catalysis. Moreover, the "return" of belt sulfurs to both M-clusters in Av1*(TOD) provides strong validation for the asymmetric displacement of the belt sulfurs in the two M-clusters of Av1*.

Our observation that the S2B, S3A, and S5A sites can each be displaced by a dinitrogen species is consistent with the previously observed "migration" of a Se reporter in the entire belt region upon turnover (*14*) and points to a possible involvement of all three sites in N_2 reduction. At the S3A site, the backbone amides of α-G356$^{(C)}$ and α-G357$^{(C)}$ that interact with the dinitrogen species are unlikely to serve as proton donors for N_2 reduction. At the S2B site, however, there is a hydrogen bond between α-H195$^{(A)}$ and the dinitrogen species that could provide protons for N_2 reduction. Additionally, an elongation of the Mo–O7 (hydroxyl) distance to 2.7 Å would be consistent with a protonation event that breaks this bond in exchange for N_2 reduction. At the S5A site, the dinitrogen species seems primed to accept protons from hydrogen bonds with α-R96$^{(C)}$ and a nearby water molecule. Moreover, an elongation of the Mo–O5 (carboxyl) distance to 2.7 Å allows O5 to participate in a hydrogen bond with a nearby water molecule that is positioned at 2.5 Å from O5 and ~4 Å from the S5A site. With ample proton sources available, the dinitrogen species at S2B and S5A are likely more protonated and/or reduced than N_2. Notably, the binding conformation of dinitrogen species at S2B is similar to the cis-$\mu_{1,2}$-binding mode of a diazene adduct to a synthetic compound (*22*, *23*). Moreover, modeling of allowable N–N bond distances suggests the possibility of having diazene-level species bound at both the S2B and S5A sites while disfavoring the presence of a hydrazine-level species at the S5A site (fig. S9). Although many mechanisms can be proposed to explain our observations (figs. S20 to S24; also see discussion in supplementary materials), experimental support is yet to be acquired for any of these proposals. The possibility that all belt-sulfur sites are involved in catalysis, stemming from our observation of asymmetric belt-sulfur displacements in the two αβ dimers of Av1, should provoke a (re)calibration in the mechanistic thinking of nitrogenase, with the ultimate goal of elucidating the intricate mechanism of enzymatic N_2 reduction.

REFERENCES AND NOTES

1. B. K. Burgess, D. J. Lowe, *Chem. Rev.* **96**, 2983–3012 (1996).
2. D. C. Rees *et al.*, *Philos. Trans. R. Soc. London Ser. A* **363**, 971–984 (2005).
3. B. M. Hoffman, D. Lukoyanov, Z. Y. Yang, D. R. Dean, L. C. Seefeldt, *Chem. Rev.* **114**, 4041–4062 (2014).
4. J. A. Wiig, Y. Hu, C. Chung Lee, M. W. Ribbe, *Science* **337**, 1672–1675 (2012).
5. T. Spatzal *et al.*, *Science* **334**, 940 (2011).
6. C. C. Lee, Y. Hu, M. W. Ribbe, *Science* **329**, 642 (2010).
7. Y. Hu, C. C. Lee, M. W. Ribbe, *Science* **333**, 753–755 (2011).
8. D. J. Lowe, R. N. Thorneley, *Biochem. J.* **224**, 877–886 (1984).
9. R. N. Thorneley, D. J. Lowe, *Biochem. J.* **224**, 887–894 (1984).
10. L. M. Cameron, B. J. Hales, *Biochemistry* **37**, 9449–9456 (1998).
11. S. C. Lee, R. H. Holm, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3595–3600 (2003).
12. T. Spatzal, K. A. Perez, O. Einsle, J. B. Howard, D. C. Rees, *Science* **345**, 1620–1623 (2014).
13. D. Sippel *et al.*, *Science* **359**, 1484–1489 (2018).
14. T. Spatzal, K. A. Perez, J. B. Howard, D. C. Rees, *eLife* **4**, e11620 (2015).
15. See supplementary materials for more details.
16. Y. Hu, A. W. Fay, B. Schmid, B. Makar, M. W. Ribbe, *J. Biol. Chem.* **281**, 30534–30541 (2006).
17. J. W. Peters *et al.*, *Biochemistry* **36**, 1181–1187 (1997).
18. C. P. Owens, F. E. Katz, C. H. Carter, V. F. Oswald, F. A. Tezcan, *J. Am. Chem. Soc.* **138**, 10124–10127 (2016).
19. A. J. Pierik, H. Wassink, H. Haaker, W. R. Hagen, *Eur. J. Biochem.* **212**, 51–61 (1993).
20. C.-H. Kim, W. E. Newton, D. R. Dean, *Biochemistry* **34**, 2798–2808 (1995).
21. P. M. Benton *et al.*, *Biochemistry* **40**, 13816–13825 (2001).
22. Y. Chen *et al.*, *J. Am. Chem. Soc.* **133**, 1147–1149 (2011).
23. Y. Li *et al.*, *Nat. Chem.* **5**, 320–326 (2013).
24. W. Kang, C. C. Lee, A. J. Jasniewski, M. W. Ribbe, Y. Hu, Anomalous datasets, Version 1.0, Zenodo (2020); https://doi.org/10.5281/zenodo.3756201.

ACKNOWLEDGMENTS

We thank D. Rees (Caltech) for insightful discussions and J. Kaiser (Molecular Observatory at Caltech) for technical assistance. **Funding:** The authors were supported by the Department of Energy grant DOE(BES) DE-SC0016510 (to Y.H. and M.W.R.), which funded work related to the mechanistic investigation of ammonia synthesis by nitrogenase and its homologs. This work was also supported by NIH-NIGMS grant GM67626 (to M.W.R. and Y.H.), which funded the crystallographic setup in the Ribbe and Hu laboratories and the development of reductant-free protein purification strategies for work related to the assembly and catalysis of nitrogenase. The authors acknowledge the Gordon and Betty Moore Foundation, the Beckman Institute, and the Sanofi-Aventis Bioengineering Research Program at Caltech for their generous support of the Molecular Observatory at Caltech. Operations at SSRL are supported by DOE and NIH. **Author contributions:** W.K. determined the structures. C.C.L. performed biochemical and spectroscopic experiments. A.J.J. and C.C.L. analyzed data with respect to chemical and mechanistic implications. M.W.R. and Y.H. designed experiments, analyzed data, and wrote the manuscript. **Competing interests:** The authors declare no competing financial or nonfinancial interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. The structural models have been deposited in the Protein Data Bank under IDs 6UG0 and 6VXT. The anomalous structure factors for 6UG0 and 6VXT are archived in Zenodo (*24*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1381/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S24
Tables S1 to S4
References (*25*–*45*)
MDAR Reproducibility Checklist

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

29 September 2019; resubmitted 11 March 2020
Accepted 20 April 2020
10.1126/science.aaz6748

PHASE SEPARATION

Partitioning of cancer therapeutics in nuclear condensates

Isaac A. Klein[1,2]*, Ann Boija[1]*, Lena K. Afeyan[1,3], Susana Wilson Hawken[1,3], Mengyang Fan[4,5], Alessandra Dall'Agnese[1], Ozgur Oksuz[1], Jonathan E. Henninger[1], Krishna Shrinivas[6,7], Benjamin R. Sabari[1], Ido Sagi[1], Victoria E. Clark[1,8], Jesse M. Platt[1,9], Mrityunjoy Kar[10], Patrick M. McCall[10,11,12], Alicia V. Zamudio[1,3], John C. Manteiga[1,3], Eliot L. Coffey[1,3], Charles H. Li[1,3], Nancy M. Hannett[1], Yang Eric Guo[1], Tim-Michael Decker[13], Tong Ihn Lee[1], Tinghu Zhang[4,5], Jing-Ke Weng[1,3], Dylan J. Taatjes[13], Arup Chakraborty[6,7,14,15,16,17,18], Phillip A. Sharp[3,18], Young Tae Chang[19], Anthony A. Hyman[11,20], Nathanael S. Gray[4,5], Richard A. Young[1,3]†

The nucleus contains diverse phase-separated condensates that compartmentalize and concentrate biomolecules with distinct physicochemical properties. Here, we investigated whether condensates concentrate small-molecule cancer therapeutics such that their pharmacodynamic properties are altered. We found that antineoplastic drugs become concentrated in specific protein condensates in vitro and that this occurs through physicochemical properties independent of the drug target. This behavior was also observed in tumor cells, where drug partitioning influenced drug activity. Altering the properties of the condensate was found to affect the concentration and activity of drugs. These results suggest that selective partitioning and concentration of small molecules within condensates contributes to drug pharmacodynamics and that further understanding of this phenomenon may facilitate advances in disease therapy.

The five to 10 billion protein molecules of cells are compartmentalized into both membrane-bound and non–membrane-bound organelles (*1*–*3*). Many non–membrane-bound organelles are phase-separated biomolecular condensates with distinct physicochemical properties that can absorb and concentrate specific proteins and nucleic acids (*4*–*17*). We reasoned that selective condensate partitioning might also occur with small-molecule drugs with targets that occur within condensates (Fig. 1A), and that the therapeutic index and efficacy of such compounds might therefore relate to their ability to partition into condensates that harbor their target. To test this idea, we focused our study on a collection of nuclear condensates previously reported in cell lines, demonstrated that they all occur in normal human cells and in tumor cells, and then developed in vitro condensate droplet assays with key components of each of the nuclear condensates to enable testing of small molecules.

Nuclear condensates have been described in diverse cultured cell lines and contain one or more proteins that can serve both as markers of the condensate and as a scaffold for condensate formation in droplet assays in vitro (*10*–*12*, *18*–*31*). Specifically, transcriptional condensates are marked by the condensate-forming proteins MED1 and BRD4 (*10*, *12*, *19*), splicing speckles by SRSF2 (*11*, *20*), heterochromatin by HP1α (*21*, *22*), and nucleoli by FIB1 and NPM1 (*23*–*25*) (fig. S1A). To determine whether such condensates can also be observed in the cells of healthy and malignant human tissue, we obtained biopsies of breast ductal epithelium, invasive ductal carcinoma, normal colon, and colon cancer (fig. S1, B and C). Immunofluorescence revealed nuclear bodies containing these marker proteins in both normal and transformed tissue (Fig. 1, B and C). There was a broad distribution of nuclear body sizes and numbers, as expected for dynamic biomolecular condensates, and no significant differences were observed between benign and malignant tissue (fig. S2, A to C). However, tumor cells acquire large superenhancers (SEs) at driver oncogenes (*32*) and these can form tumor-specific transcriptional condensates.

We developed an assay to model these nuclear condensates and study the behavior of small molecules within these droplets (Fig. 1D). We produced and purified recombinant, fluorescently labeled versions of MED1, BRD4, SRSF2, HP1α, FIB1, and NPM1 (fig. S3, A and B) and confirmed the ability of these proteins to form droplets in an in vitro assay (fig. S4, A and B). To investigate the partitioning behavior of small molecules, we added the dyes fluorescein (332 Da) and Hoechst (452 Da), as well as fluorescently labeled dextrans (4.4 kDa), to solutions containing each of the six protein condensates. The dyes and dextrans appeared to diffuse through all the condensates without substantial partitioning (Fig. 1E and figs. S5 and S6, A to D). Small-molecule drugs are generally smaller than 1 kDa, so these results suggested that small molecules can freely diffuse through these nuclear condensates unless there are factors other than size that influence partitioning.

We next sought to determine whether diverse clinically important drugs with targets that reside in nuclear condensates also exhibit free diffusion across these condensates or if they display a different behavior. Cisplatin and mitoxantrone, members of a class of antineoplastic compounds that modify DNA through platination or intercalation, can either be modified to have fluorescent properties (cisplatin) (*33*) or are already inherently fluorescent (mitoxantrone). When added to droplet formation buffer with purified MED1, BRD4, SRSF2, HP1α, FIB1, or NPM1, cisplatin was found to be selectively concentrated in MED1 droplets (Fig. 2A and fig. S7A), with a partition coefficient of up to 600 (fig. S8, A to C). Fluorescent modification of cisplatin did not appear to contribute to this behavior in vitro, because the modified drug could be chased out of the condensate with unmodified cisplatin, and an isomer of cisplatin did not exhibit the same behavior (fig. S7, B to D). Mitoxantrone was also concentrated in MED1 condensates, as well as in FIB1 and NPM1 condensates (Fig. 2B and fig. S7A). Consistent with these results, mitoxantrone is known to concentrate in the nucleolus, where FIB1 and NPM1 reside (*34*, *35*). These results show that, in contrast to the dyes tested above, small-molecule drugs may concentrate in certain condensates even in the absence of the drug target.

We selected for further study antineoplastic drugs that target transcriptional regulators

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. [2]Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. [5]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [6]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [7]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [8]Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA. [9]Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital, Boston, MA, 02114, USA. [10]Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany. [11]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. [12]Center for Systems Biology Dresden, 01307 Dresden, Germany. [13]Department of Biochemistry, University of Colorado, Boulder, CO 80303, USA. [14]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [15]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [16]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [17]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard Medical School, Cambridge, MA 02139, USA. [18]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [19]Department of Chemistry, Pohang University of Science and Technology, and Center for Self-assembly and Complexity, Institute for Basic Science (IBS), Pohang 37673, Republic of Korea. [20]Cluster of Excellence Physics of Life, Technical University of Dresden, 01062 Dresden, Germany.
*These authors contributed equally to this work.
†Corresponding author. Email: young@wi.mit.edu

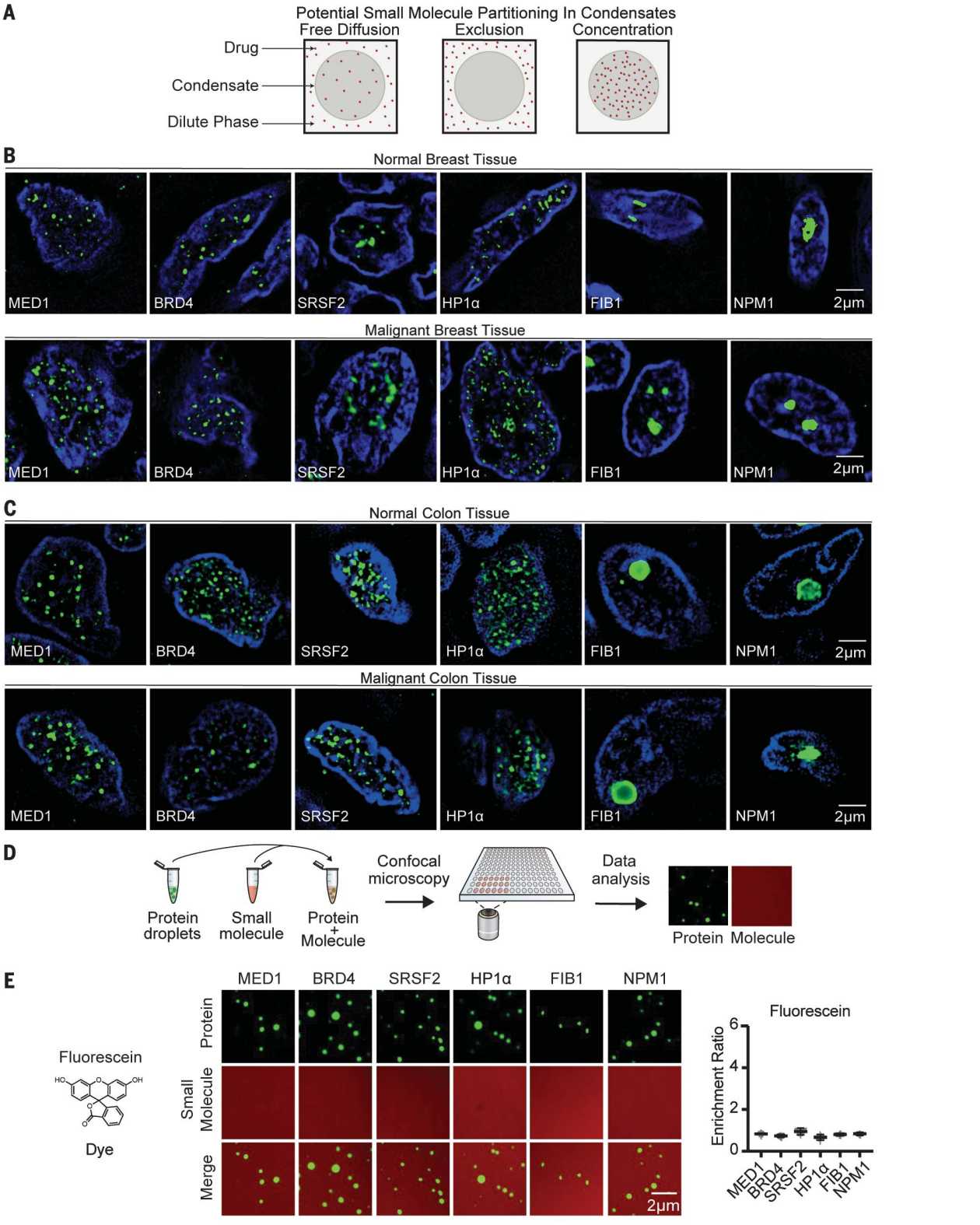


Fig. 1. Nuclear condensates in human tissue and in vitro. (**A**) Model illustrating the potential behaviors of small molecules in nuclear condensates. (**B** and **C**) Immunofluorescence of scaffold proteins of various nuclear condensates in tissue biopsies from benign and malignant human breast tissue (B) and from benign and malignant colon tissue (C) in nuclei stained with Hoechst and imaged at 100× on a fluorescent confocal microscope (see also figs. S1 and S2). (**D**) Schematic of in vitro droplet formation assay to measure small-molecule partitioning into nuclear condensates. (**E**) In vitro droplet assay showing the behavior of fluorescein dye in the presence of six protein condensates formed in 125 mM NaCl and 10% PEG with 10 μM protein and 5 μM fluorescein imaged at 150× on a confocal fluorescent microscope (see also figs. S3 to S6). Quantification of enrichment of the drug is shown on the right. Error bars represent SEM.

expected to be contained within transcriptional condensates in cells. These targets include: (i) the estrogen receptor (ER), a transcription factor and nuclear hormone receptor; (ii) CDK7, a cyclin-dependent kinase that functions in transcription initiation and cell cycle control; and (iii) BRD4, a bromodomain protein and coactivator involved in oncogene regulation (fig. S9, A and B). To monitor drug behavior with a confocal fluorescent microscope, we used a fluorescent tamoxifen analog (FLTX1) that targets ER and modified fluorescent THZ1 and JQ1, which target CDK7 and BRD4, respectively (*36*, *37*). FLTX1 and THZ1 concentrated preferentially in MED1 droplets (Fig. 2, C and D, and fig. S7A), and this behavior was not attributable to the fluorescent moiety (fig. S7, B and D). JQ1 concentration presented a different pattern, being concentrated in MED1, BRD4, and NPM1 droplets (Fig. 2E and fig. S7, A and B). Reinforcing these results, we found that the small molecules that concentrated in MED1 condensates were also concentrated in condensates formed from purified whole Mediator complexes (fig. S10A) and in MED1 condensates formed in an alternative crowding agent (fig. S11A). The targets of these three compounds (ERα, CDK7, and the bromodomains of BRD4) are not present in these in vitro condensates but are present in the SEs that form condensates with transcription factors and Mediator in vivo (*10*, *12*, *38*) (fig. S9, A and B), suggesting that the ability of some small molecules to concentrate preferentially in the same condensate as their protein target may contribute to the pharmacological properties of these drugs.

To gain additional insight into the nature of interactions governing small-molecule enrichment in condensates, we focused on the MED1-IDR condensate. Fluorescence recovery after photobleaching (FRAP) experiments showed that cisplatin molecules were highly mobile in this condensate (fig. S12, A and B), suggesting that the condensate produces a physicochemical environment that facilitates drug concentration in a state of high dynamic mobility. To gain insights into the chemical features of small molecules that may contribute to selective association with MED1 in condensates, we used a fluorescent boron-dipyrromethene (BODIPY) library of 81 compounds with various combinations of chemical side groups (fig. S13A). Molecules that contained aromatic rings were found to preferentially concentrate in MED1 condensates (figs. S13, A to D, and S14A). These data suggest that pi–pi or pi–cation interactions are among the physicochemical properties that favor small-molecule partitioning into MED1 condensates. Aromatic amino acids participate in pi–system interactions and are overrepresented in the MED1 IDR relative to the other condensate-forming proteins studied (fig. S3B). We generated a MED1 aromatic mutant protein (all 30 aromatic amino acids mutated to alanine) that retained the ability to form droplets in vitro, indicating that the aromatic amino acids are not required for droplet formation (fig. S14, B and C), but small-molecule probes containing aromatic rings and the polar molecule cisplatin no longer partitioned into condensates

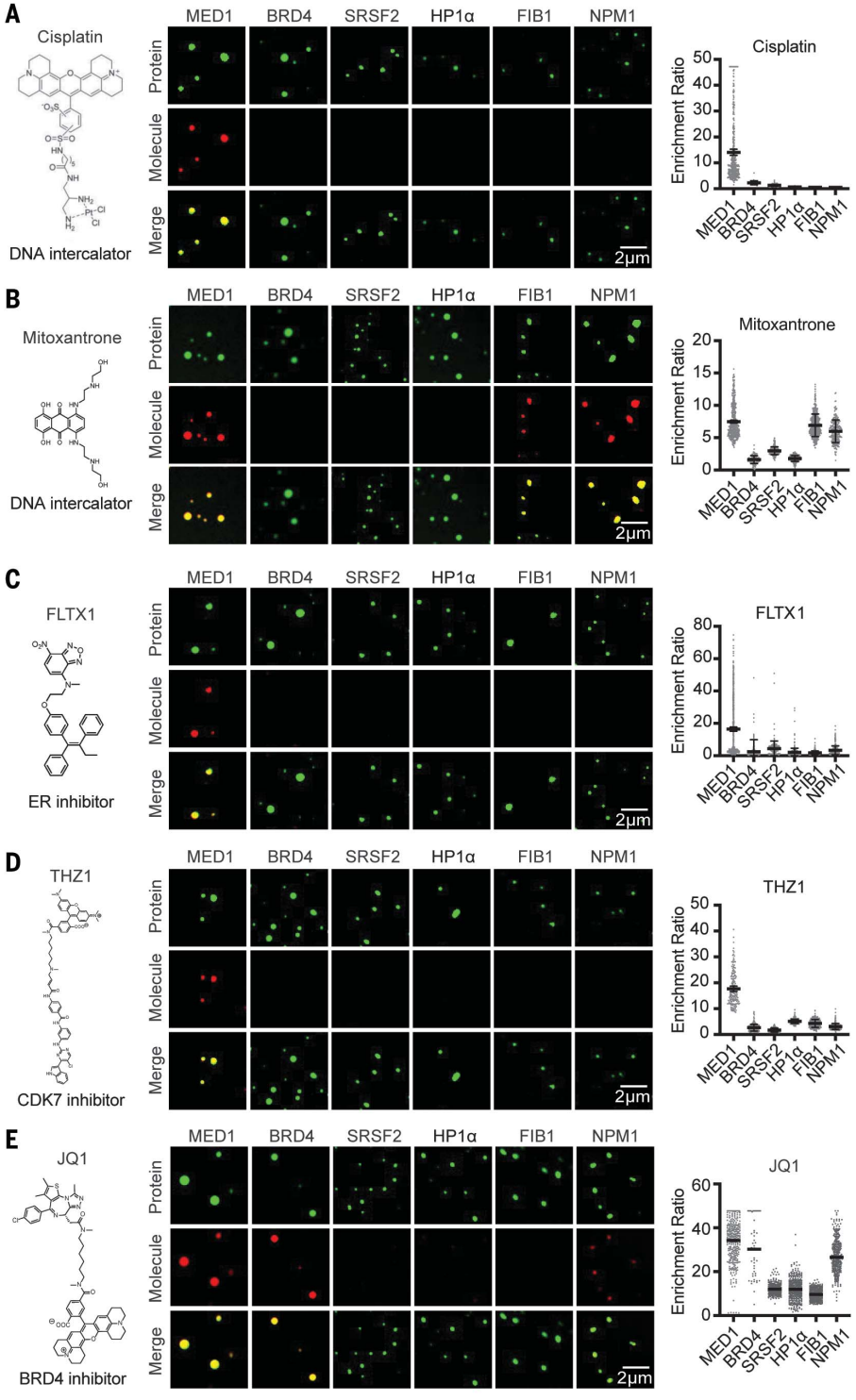


Fig. 2. Partitioning behavior of small-molecule drugs in nuclear condensates in a droplet assay. Six nuclear condensates formed in 125 mM NaCl and 10% PEG with 10 μM protein treated with (**A**) 5μM cisplatin-TMR, (**B**) 50 μM mitoxantrone, (**C**) 100 μM FLTX1, (**D**) 5 μM THZ1-TMR, or (**E**) 1 μM JQ1-ROX imaged at 150× on a confocal fluorescent microscope (see also figs. S7 to S11). Quantification of enrichment of the drug within droplets is shown on the right of each panel. Error bars represent SEM (see also figs. S12 to S14).

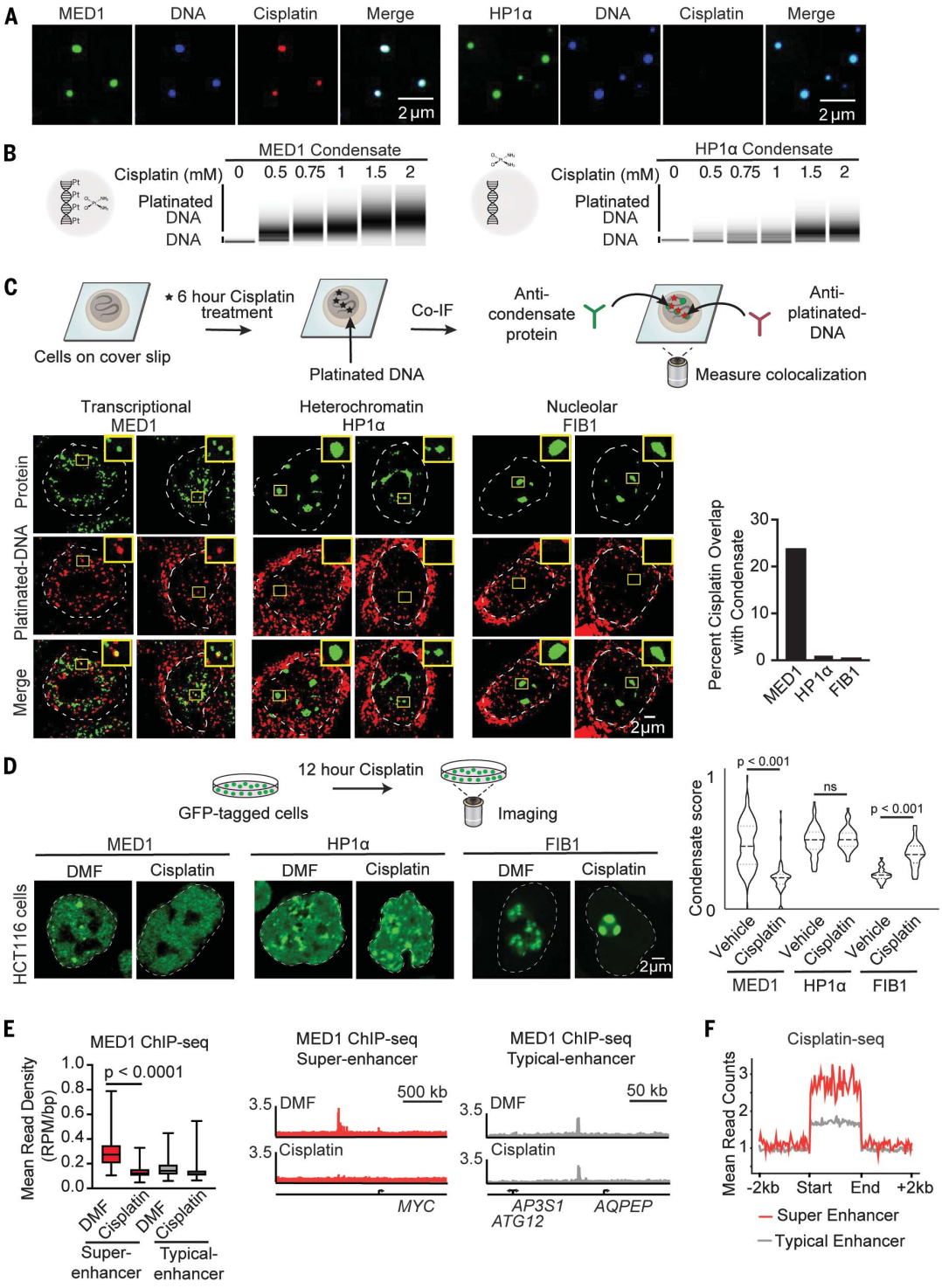


Fig. 3. Small-molecule concentration within condensates influences drug activity. (**A**) In vitro droplet assay of MED1 and HP1α condensates formed in 125 mM NaCl and 10% PEG, 5 nM of 450-bp DNA, 10 μM MED1, and 5 μM cisplatin-TR imaged at 150× on a confocal fluorescent microscope (see also fig. S15). (**B**) Bioanalyzer tracings of DNA contained within either MED1 or HP1α droplets exposed to the indicated concentration of cisplatin. (**C**) (Top) Schematic of an assay to determine the location of platinated DNA relative to various nuclear condensates. (Bottom) Coimmunofluorescence of platinated DNA and the indicated protein in HCT116 cells treated with 50 μM cisplatin for 6 hours imaged at 100× on a confocal fluorescent microscope. Quantification of overlap is shown on the right. (**D**) (Top) Schematic of a live-cell condensate dissolution assay. (Bottom) HCT116 cells bearing endogenously mEGFP-tagged MED1, HP1α, or FIB1 treated with 50 μM cisplatin for 12 hours. Quantification of MED1, HP1α, or FIB1 condensate score is shown on the right. (**E**) MED1 ChIP-seq in HCT116 cells treated with vehicle or 50 μM cisplatin for 6 hours. (Left) Mean read density of MED1 at SEs and typical enhancers. Error bars represent minimum and maximum. (Right) Gene tracks of MED1 ChIP-Seq at the *MYC* SE and AQPEP typical enhancer. (**F**) Metaplot of cisplatin-DNA-Seq in cisplatin-treated HeLa cells comparing SEs and typical enhancers (*40*) (see also figs. S16 to S21).

formed by the MED1 aromatic mutant protein (fig. S14, D to F). These results suggest that the aromatic residues of MED1 condensates contribute to the physicochemical properties that selectively concentrate these small molecules.

We anticipated that the ability of small molecules to concentrate in specific condensates would influence target engagement and thus drug pharmacodynamics. To investigate this, we took advantage of the ability of condensates to incorporate DNA (Fig. 3A and fig. S15A) and measured the relative efficiency of DNA platination by cisplatin in MED1 condensates, where cisplatin is concentrated, versus HP1α condensates, where cisplatin freely diffuses (Fig. 2A). DNA platination, visualized by size shift on a bioanalyzer, was more prevalent in MED1 condensates than in HP1α condensates (Fig. 3B), consistent with the expectation that elevated concentrations of cisplatin in the MED1 condensates yield enhanced target

engagement. If cisplatin becomes concentrated in Mediator condensates in cells, then we would expect that DNA colocalized within Mediator condensates would be preferentially platinated. To test this idea, we performed coimmunofluorescence in cisplatin-treated HCT116 colon cancer cells using an antibody that specifically recognizes platinated DNA (fig. S16A) (*39*), together with antibodies specific for MED1, HP1α, or FIB1. Consistent with cisplatin's preference for MED1 condensates in vitro, we found that platinated DNA frequently colocalized with MED1 condensates but not with HP1α or FIB1 condensates (Fig. 3C). To determine whether the ability of cisplatin to engage DNA is dependent on the presence of a MED1 condensate, we treated cells with JQ1, which caused a loss of MED1 condensates (fig. S16B), and observed a concomitant reduction in platinated DNA at the *MYC* oncogene (fig. S16, C and D). These results are consistent with the idea that the concentration of small molecules in specific condensates can influence the efficiency of target engagement.

In cells, the preferential modification of DNA in MED1-containing condensates might be expected to selectively disrupt these condensates with prolonged treatment. To test this, HCT116 colon cancer cells were engineered

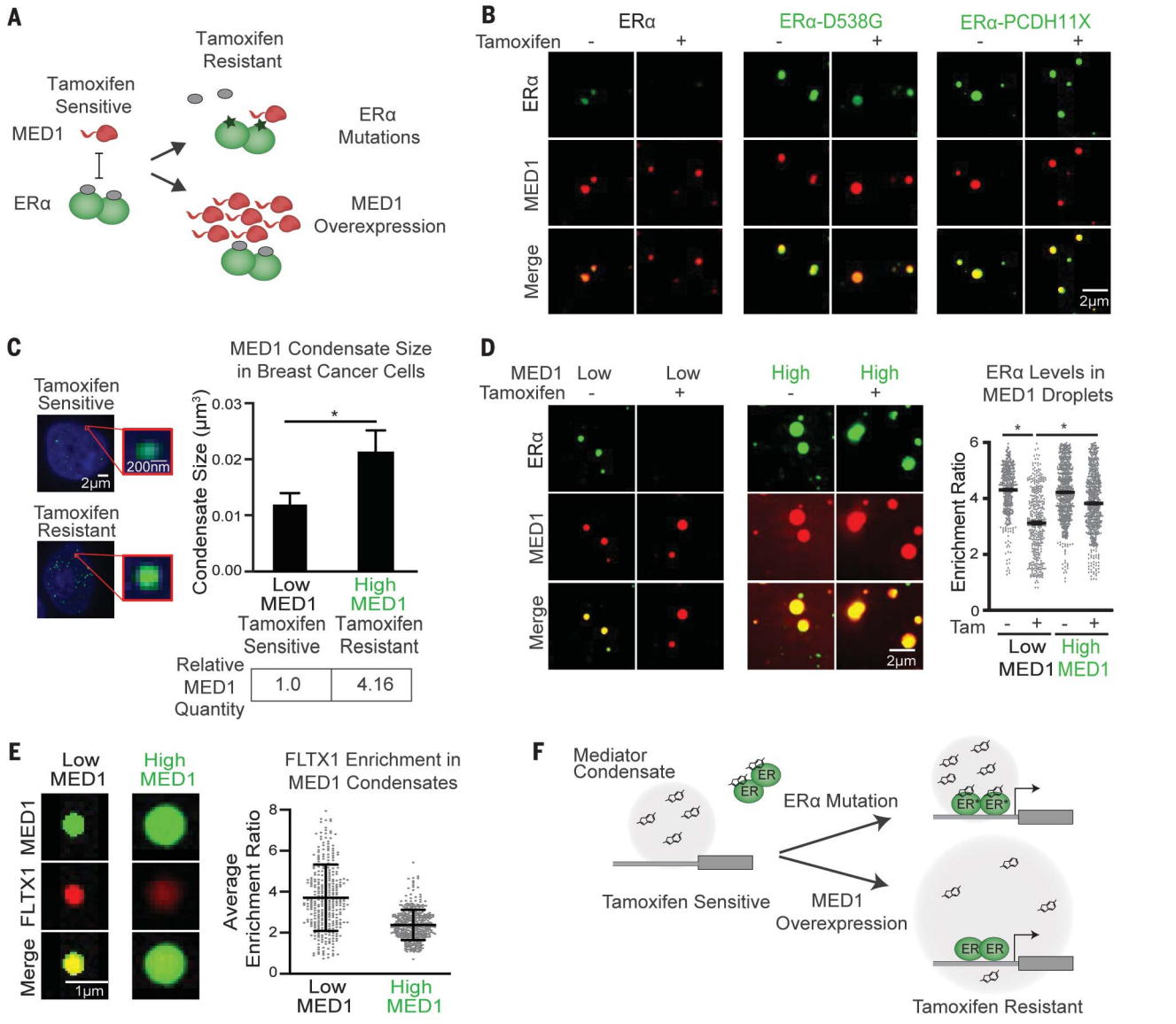


Fig. 4. Tamoxifen action and resistance in MED1 condensates. (**A**) Schematic showing tamoxifen resistance by ER mutation and MED1 overexpression in breast cancer. (**B**) In vitro droplet assay of the indicated form of GFP-tagged ER in the presence of estrogen with and without 100 μM tamoxifen. Droplets are formed in 125 mM NaCl and 10% PEG with a 10 μM concentration of each protein and 100 μM estrogen. (**C**) (Left) Immunofluorescence of MED1 in tamoxifen-sensitive (MCF7) and tamoxifen-resistant (TAMR7) ER$^+$ breast cancer cell lines imaged at 100× on a confocal fluorescent microscope. (Top right) Quantification of MED1 condensate size in breast cancer cells. (Bottom right) Relative quantities of MED1 in the indicated breast cancer cell line by Western blot. Error bars represent SEM. (**D**) In vitro droplet assays of ER in the presence of 100 μM estrogen with and without 100 μM tamoxifen with either 5 μM (low) or 20 μM (high) MED1. Droplets are formed with 5 μM ER in 125 mM NaCl and 10% PEG and imaged at 150× on a confocal fluorescent microscope. Error bars represent SEM. (**E**) In vitro droplet assay with either 5 μM (low) or 20 μM (high) MED1 with 100 μM FLTX1 in 125 mM NaCl and 10% PEG. Error bars represent SD. (**F**) Models for tamoxifen resistance caused by altered drug affinity (through ER mutation) or concentration (through MED1 overexpression) (see also figs. S22 to S30).

to express green fluorescent protein (GFP)–tagged marker proteins for each of the six nuclear condensates (figs. S17, A to F, and S18, A and B). When exposed to cisplatin, a selective and progressive reduction in MED1 condensates was observed (Fig. 3D and figs. S19, A and B, and S20A). Consistent with this, cisplatin treatment led to a preferential loss of MED1 chromatin immunoprecipitation sequencing (ChIP-seq) signal at SEs (Fig. 3E and fig. S21A). Furthermore, high-throughput sequencing data from platinated-DNA pulldown (*40*) revealed that cisplatin-modified DNA preferentially occurs at SEs, where MED1 is concentrated (*41*) (Fig. 3F). These results are consistent with reports that cisplatin preferentially modifies transcribed genes (*40*, *42*) and suggest that this effect is due to preferential condensate partitioning. Taken together, these results suggest a model in which cisplatin preferentially modifies SE DNA, which in turn leads to dissolution of these condensates. Previous studies have shown that diverse tumor cells become highly dependent on SE-driven oncogene expression (*43*–*47*), which might explain why platinum drugs, which are capable of general DNA modification, are effective therapeutics in diverse cancers (*48*).

We explored the behavior of another clinically important antineoplastic drug, tamoxifen, to assess whether its drug response and resistance were associated with partitioning in condensates (Fig. 4A). ERα incorporates into MED1 condensates in an estrogen-dependent manner in vitro (*12*); droplet assays confirmed this and revealed that the addition of tamoxifen leads to eviction of ERα from the MED1 condensates (Fig. 4B). We further investigated the effects of estrogen and tamoxifen on MED1 condensates in breast cancer cells, focusing on the *MYC* oncogene because of its prominent oncogenic role and responsiveness to estrogen (*49*). MED1 condensates were observed on the *MYC* oncogene in the ER^+ breast cancer cell line MCF7 (figs. S9A and S22, A to D). DNA fluorescence in situ hybridization with MED1 immunofluorescence revealed that estrogen enhances the formation of MED1 condensates at the *MYC* oncogene and that tamoxifen treatment reduces it (fig. S23, A and B). Artificial MED1 condensates without ER concentrated FLTX1 at the site of the condensate (fig. S24A), indicating that ER is not required for the partitioning of FLTX1 into MED1 condensates in cells. These results are consistent with the model that ERα interacts with MED1 condensates in an estrogen-dependent, tamoxifen-sensitive manner to drive oncogene expression in breast cancer cells.

The mechanisms that produce drug resistance can provide clues to drug activity in the clinical setting. Endocrine therapy and tamoxifen resistance are an enduring clinical challenge and are associated with multiple mechanisms, including ERα mutation and MED1 overexpression (Fig. 4A and fig. S25) (*50*–*54*). To investigate whether ERα mutations alter ERα behavior in condensates, we produced four patient-derived ERα mutant proteins and tested their partitioning in the presence of tamoxifen. In contrast to wild-type ERα, condensates composed of patient-derived ERα mutants and MED1 were not disrupted upon tamoxifen treatment (Fig. 4B and fig. S26, A and B). The ERα point mutations reduce the affinity for tamoxifen by ~10-fold (*51*), indicating that the drug concentration in the droplet is inadequate to evict these ER mutant proteins when this affinity is reduced.

MED1 overexpression is associated with tamoxifen resistance and poor prognosis in breast cancer (*50*), but it is not clear why overexpression of one subunit of the Mediator complex produces resistance. We considered the possibility that overexpressed MED1 is incorporated into transcriptional condensates, which contain clusters of Mediator molecules (*38*), thereby expanding their volumes and diluting the available tamoxifen (fig. S27A). We found that the tamoxifen-resistant breast cancer cell line TAMR7 (*55*), which was derived from the tamoxifen-sensitive cell line MCF7, produced fourfold elevated levels of the MED1 protein (fig. S27B). The volume of MED1-containing condensates was twofold larger in these cells (Fig. 4C and fig. S27C). When modeled in an in vitro droplet assay, we found that a fourfold increase in MED1 levels led to a commensurate increase in droplet size (fig. S28, A and B). Furthermore, we found that 100 μM tamoxifen prevented ERα incorporation into MED1 condensates (Fig. 4, B and D), but was much less effective in preventing ERα incorporation into the larger MED1 condensates produced with higher MED1 levels (Fig. 4D). To confirm that the levels of tamoxifen in the larger droplets are more dilute, we measured the enrichment of the fluorescent tamoxifen analog FLTX1 in MED1 droplets and found that the larger condensates had lower concentrations of the drug (Fig. 4E). These results were mirrored in cells, where a collection of tethered ERα molecules formed a MED1 condensate that was eliminated by tamoxifen, but when MED1 was overexpressed, tamoxifen was unable to dissociate the ERα-MED1 condensate (fig. S29A). Similarly, knockdown of MED1 in tamoxifen-resistant breast cancer cells sensitizes cells to tamoxifen (*50*, *54*). These results support a model of tamoxifen resistance where MED1 overexpression causes the formation of larger transcriptional condensates in which tamoxifen is diluted and is thereby less effective at dissociating ER from the condensate (Fig. 4F).

Our results show that drugs partition selectively into condensates, that this can occur through physicochemical properties that exist independently of their molecular targets, and that cells can develop resistance to drugs through condensate-altering mechanisms. This may explain the surprising observation that inhibition of global gene regulators such as BRD4 or CDK7 can have selective effects on oncogenes that have acquired large SEs (*45*); selective partitioning of inhibitors such as JQ1 and THZ1 into SE condensates will preferentially disrupt transcription at those loci. These results also have implications for the future development of efficacious disease therapeutics; effective target engagement will depend on measurable factors such as drug partitioning in condensates (fig. S30, A to D). Condensate assays of the type described here may thus help to optimize condensate partitioning, target engagement, and the therapeutic index of small-molecule drugs.

REFERENCES AND NOTES

1. Y. Shin, C. P. Brangwynne, *Science* **357**, eaaf4382 (2017).
2. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
3. A. A. Hyman, C. A. Weber, F. Jülicher, *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
4. R. J. Ries *et al.*, *Nature* **571**, 424–428 (2019).
5. E. M. Langdon, A. S. Gladfelter, *Annu. Rev. Microbiol.* **72**, 255–271 (2018).
6. E. M. Langdon *et al.*, *Science* **360**, 922–927 (2018).
7. M.-T. Wei, Y.-C. Chang, S. F. Shimobayashi, Y. Shin, C. P. Brangwynne, Nucleated transcriptional condensates amplify gene expression. bioRxiv 737387 [Preprint]. 21 August 2019; https://doi.org/10.1101/737387.
8. T. J. J. Nott *et al.*, *Mol. Cell* **57**, 936–947 (2015).
9. T. J. Nott, T. D. Craggs, A. J. Baldwin, *Nat. Chem.* **8**, 569–575 (2016).
10. B. R. Sabari *et al.*, *Science* **361**, eaar3958 (2018).
11. Y. E. Guo *et al.*, *Nature* **572**, 543–548 (2019).
12. A. Boija *et al.*, *Cell* **175**, 1842–1855.e16 (2018).
13. J. J. Bouchard *et al.*, *Mol. Cell* **72**, 19–36.e8 (2018).
14. K. Shrinivas *et al.*, *Mol. Cell* **75**, 549–561.e7 (2019).
15. E. P. Bentley, B. B. Frey, A. A. Deniz, *Chemistry* **25**, 5600–5610 (2019).
16. M. Boehning *et al.*, *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
17. P. Cramer, *Nature* **573**, 45–54 (2019).
18. S. Alberti *et al.*, *J. Mol. Biol.* **430**, 4806–4820 (2018).
19. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, *Cell* **169**, 13–23 (2017).
20. Y. Chen, A. S. Belmont, *Curr. Opin. Genet. Dev.* **55**, 91–99 (2019).
21. A. G. Larson *et al.*, *Nature* **547**, 236–240 (2017).
22. A. R. Strom *et al.*, *Nature* **547**, 241–245 (2017).
23. M. Feric *et al.*, *Cell* **165**, 1686–1697 (2016).
24. D. M. Mitrea *et al.*, *eLife* **5**, e13571 (2016).
25. D. M. Mitrea *et al.*, *Nat. Commun.* **9**, 842 (2018).
26. S. F. Banani *et al.*, *Cell* **166**, 651–663 (2016).
27. A. Patel *et al.*, *Cell* **162**, 1066–1077 (2015).
28. S. Alberti, *J. Cell Sci.* **130**, 2789–2796 (2017).
29. A. E. Posey, A. S. Holehouse, R. V. Pappu, "Chapter one - phase separation of intrinsically disordered proteins," in *Methods in Enzymology* (Elsevier, 2018), vol. 611, pp. 1–30.
30. V. N. Uversky, *Curr. Opin. Struct. Biol.* **44**, 18–30 (2017).
31. Y. S. Mao, B. Zhang, D. L. Spector, *Trends Genet.* **27**, 295–306 (2011).
32. D. Hnisz *et al.*, *Cell* **155**, 934–947 (2013).
33. Y. H. Chu *et al.*, *Mol. Pharm.* **13**, 2677–2682 (2016).
34. S. Vibet *et al.*, *Drug Metab. Dispos.* **35**, 822–828 (2007).
35. P. J. Smith, H. R. Sykes, M. E. Fox, I. J. Furlong, *Cancer Res.* **52**, 4000–4008 (1992).
36. N. Kwiatkowski *et al.*, *Nature* **511**, 616–620 (2014).
37. J. Marrero-Alonso *et al.*, *Eur. J. Pharm. Biopharm.* **85** (3 Pt B), 898–910 (2013).
38. W.-K. Cho *et al.*, *Science* **361**, 412–415 (2018).

39. M. J. Tilby *et al*., *Cancer Res.* **51**, 123–129 (1991).
40. X. Shu, X. Xiong, J. Song, C. He, C. Yi, *Angew. Chem. Int. Ed.* **55**, 14246–14249 (2016).
41. W. A. Whyte *et al*., *Cell* **153**, 307–319 (2013).
42. X. Rovira-Clave *et al*., Subcellular localization of drug distribution by super-resolution ion beam imaging. bioRxiv 557603 [Preprint]. 22 February 2019; https://doi.org/10.1101/557603.
43. Y. Wang *et al*., *Cell* **163**, 174–186 (2015).
44. M. R. Mansour *et al*., *Science* **346**, 1373–1377 (2014).
45. J. E. Bradner, D. Hnisz, R. A. Young, *Cell* **168**, 629–643 (2017).
46. J. Lovén *et al*., *Cell* **153**, 320–334 (2013).
47. Z. Nie *et al*., *Cell* **151**, 68–79 (2012).
48. S. Dasari, P. B. Tchounwou, *Eur. J. Pharmacol.* **740**, 364–378 (2014).
49. D. Dubik, T. C. Dembinski, R. P. C. Shiu, *Cancer Res.* **47**, 6517–6521 (1987).
50. A. Nagalingam *et al*., *Carcinogenesis* **33**, 918–930 (2012).
51. S. W. Fanning *et al*., *eLife* **5**, e12792 (2016).
52. C. S. Ross-Innes *et al*., *Nature* **481**, 389–393 (2012).
53. M. Murtaza *et al*., *Nature* **497**, 108–112 (2013).
54. J. Cui *et al*., *Cancer Res.* **72**, 5625–5634 (2012).
55. S. Hole *et al*., *Int. J. Oncol.* **46**, 1481–1490 (2015).

ACKNOWLEDGMENTS

We thank C. Glinkerman for helpful comments; W. Salmon of the W.M Keck Microscopy Facility and T. Volkert, J. Love, S. Mraz, and S. Gupta of the Whitehead Genome Technologies Core for technical assistance; and staff of the light microscopy facility at the MPI-CBG in Dresden for extensive help and support. **Funding:** This work was supported by NIH grants GM123511, CA213333, and CA155258 (R.A.Y.), NSF grant PHY1743900 (R.A.Y.), funds from Novo Nordisk (R.A.Y. and P.A.S.) NIH grant GM117370 (D.J.T.), Max Planck Society (A.A.H.), American Society of Clinical Oncology Young Investigator Award (I.A.K.), American Cancer Society Postdoctoral Fellowship (I.A.K.), Ovarian Cancer Research Alliance Mentored Investigator Award (I.A.K.), Swedish Research Council Postdoctoral Fellowship (VR 2017-00372) (A.B.), Hope Funds for Cancer Research (AD), Gruss-Lipper Postdoctoral Fellowship and by the Rothschild Postdoctoral Fellowship (I.S.), NIH grant T32:5T32DK007191-45 (J.M.P.), German Research Foundation DFG postdoctoral fellowship DE 3069/1-1 (T.-M.D.), Cancer Research Institute Irvington Fellowship (Y.E.G.), Damon Runyon Cancer Research Foundation Fellowship (2309-17) (BRS), ELBE postdoctoral fellowship (P.M.). **Author contributions:** Conceptualization: I.A.K., A.B., R.A.Y.; Data curation: S.W.H., J.E.H.; Formal analysis: I.A.K., A.B., L.K.H., S.W.H., J.H., P.M.M., M.K., C.H.L.; Funding acquisition: D.J.T., A.C., P.A.S., Y.T.C., A.A.H., N.S.G., R.A.Y.; Investigation: I.A.K., A.B., L.K.A., S.W.H., M.F., A.D., O.O., K.S., B.R.S., I.S., VEC, J.M.P., M.K., P.M.M., A.V.Z., Y.E.G.; Methodology: I.A.K., A.B., L.K.A., A.D., O.O.; Project administration: I.A.K., A.B., T.I.L., R.A.Y.; Resources: M.F., J.C.M., E.L.C., C.H.L., N.M.H., T.M.D., TZ, D.J.T., A.C., Y.T.C., A.A.H., N.S.G., R.A.Y.; Software: J.E.H., K.S.; Supervision: R.A.Y.; Validation: I.A.K., A.B., L.K.A., S.W.H.; Visualization: I.A.K., A.B., L.K.A., S.W.H., A.D., J.E.H., K.S., B.R.S., M.K., P.M.M., A.Z.; Writing – original draft: I.A.K., A.B., R.A.Y.; Writing – review & editing: all authors. **Competing interests:** R.A.Y. is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics and Dewpoint Therapeutics. A.A.H. is a founder and shareholder in Dewpoint Therapeutics and a shareholder of Careway Therapeutics. N.S.G. is a founder and shareholder of Syros Pharmaceuticals. I.A.K. is a shareholder and member of the Scientific Advisory Board of Dewpoint Therapeutics. J.K.W. is a cofounder, member of the Scientific Advisory Board, and shareholder of DoubleRainbow Biosciences. T.I.L. is a shareholder of Syros Pharmaceuticals and a consultant of Camp4 Therapeutics. A.C. is on the Scientific Advisory Board of Dewpoint Therapeutics and Omega Therapeutics. P.A.S. is a shareholder and consultant of Dewpoint Therapeutics. Y.T.C. is an inventor on U.S. Patent US9513294 B2, China Patent ZL 201380067802.8, EP 2,938,619 B1, and Japan Patent 6300380 held by POSTECH University that cover diversity-oriented fluorescent library approaches. I.A.K., A.B., and R.A.Y. are inventors on patent application submitted by The Whitehead Institute that covers small molecule drug partitioning in and acting upon biomolecular condensates. All other authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. High-throughput sequencing data sets are available in GEO (GSE149085).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1386/suppl/DC1
Materials and Methods
Figs. S1 to S30
References (*56*–*68*)
MDAR Reproducibility Checklist

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

9 September 2019; resubmitted 24 February 2020
Accepted 29 April 2020
10.1126/science.aaz4427

By **Akira Nishii**

Undergrads in charge

I entered the main conference room in our chemistry building, feeling nervous. I was the last to arrive, and I hastily took my place at a boat-shaped table alongside four other students, all undergrads like me. A group of professors sat across from us, chatting amicably among themselves. Scheduling this meeting—and getting more than one faculty member in a room at once—had taken me almost a month. Determined to make the most of their time, I collected my thoughts and skimmed the meeting agenda for a fifth time. Our group's goal, simple but unprecedented, was spelled out in bold at the top of my notes: "We want to build a university course from the bottom up."

We didn't know what would come out of the meeting. We wanted to create a course because we believed the introductory chemistry curriculum did not sufficiently prepare students for work in a research lab. But we worried the professors would see us as naïve children. What did we know about designing a course? Who were we to insist that the curriculum was inadequate?

I was in year two of my bachelor's degree, but my experience on campus went back further than that. I grew up near the university and had volunteered as a research assistant during my last 2 years of high school. A postdoc in a diabetes lab had taken me under his wing, teaching me how to extract DNA, perform Western blots, culture cells, and handle mice. I was grateful for the time he spent with me, sometimes doubting whether my presence amounted to more than a burden.

When I graduated, I decided to attend the university to study biology and chemical engineering. Fortunately, my earlier lab experience helped me secure another research assistant position, this time in a cancer lab. I went into the experience feeling more confident with my lab skills, which allowed me to focus on understanding the science.

At the same time, I observed that many of my peers were at a disadvantage because of their lack of research experience. Even after joining a lab, many of them had to spend months improving their lab techniques before they were able to contribute intellectually to research projects. That led me to brainstorm possible ways to help them.

I recruited other students who were interested in chemistry research, and we set out to convince faculty members that a new course was needed to address this issue. Our university's introductory chemistry course had a lab component, but it only covered a few lab techniques and wasn't tailored to students who wanted to get involved in chemistry research.

"Students should be advocates for their own education."

The new course that we had in mind would expose younger students to a diverse range of research skills, such as reading the literature and applying standard chemistry techniques. We didn't plan to ask faculty members to teach it; instead, we thought the course could be taught by older, more experienced undergraduate students.

When we met with the professors to discuss our proposed course, we assumed they would be skeptical. We were surprised to find they were instead open to our idea. One senior professor had to leave early, but before he did, he rearticulated our position and gave me a knowing wink. A younger professor then suggested that he could provide faculty oversight, and we started to discuss how we might secure funding and departmental resources.

Our course was added to the official course catalog for the following semester, and it was extremely popular, quickly exceeding the maximum enrollment capacity of 80 students. We designed the course ourselves, and two undergraduate students—both of whom were paid by the university—served as instructors. At the end of the semester, students and professors alike were happy with the outcome, so the course has continued to be offered in the years since.

It is all too common for undergraduate students to have little say in shaping their education. I'm grateful that the professors in our department were willing to listen and help us take the curriculum in a new direction, but I suspect this level of openness is rare. Students should be advocates for their own education—and universities should welcome this ambition with open arms. ■

Akira Nishii is an undergraduate student at the University of Michigan, Ann Arbor. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS

Science Webinars

AAAS

Science Webinars help you keep pace with emerging scientific fields!

Stay informed about scientific breakthroughs and discoveries.

Gain insights into current research from top scientists.

Take the opportunity to ask questions during live broadcasts.

Get alerts about upcoming free webinars.

Sign up at: webinar.sciencemag.org/stayinformed

Universität Konstanz

The University of Konstanz is one of Germany's Universities of Excellence: It was successful in all three funding lines of the German Excellence Initiative and has also been competing successfully in both funding lines of its successor programme, the Excellence Strategy, since 2018.

The *Zukunftskolleg* is a central research institution of the University of Konstanz for early career researchers in the natural sciences, humanities and social sciences. It is an integral component of the university's strategy „creative.together". The *Zukunftskolleg* promotes interdisciplinary discourse between distinguished researchers from Germany and abroad, and provides them with the resources to conduct their research independently and free from administrative constraints. The working language is English.

The *Zukunftskolleg* of the University of Konstanz offers:

three 5-year Research Fellowships (any discipline represented at the University of Konstanz)

(Salary Scale TV-L 14)

to develop and implement individual research projects and build up their own research group.

Fellowships will begin on May 1, 2021.

Reference number 2020/102

three 2-year Postdoctoral Fellowships (any discipline represented at the University of Konstanz)

(Salary Scale TV-L 13)

to develop and implement individual research projects. Conditionally on the submission of an external grant, the position can be extended for an additional year.

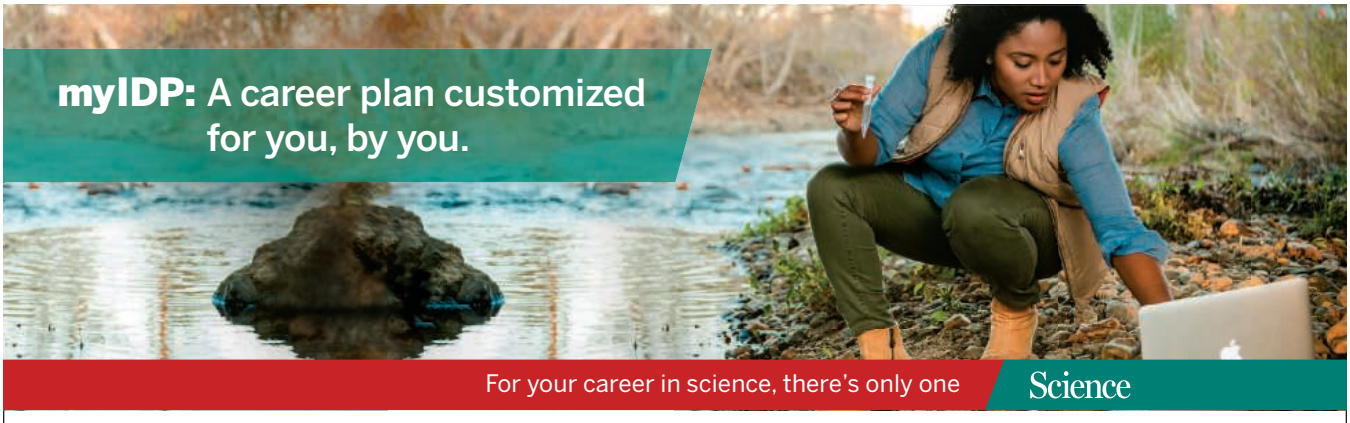
Fellowships will begin on April 1, 2021.

Reference number 2020/103

Zukunftskolleg

Applications, supporting materials and two letters of reference should be submitted in English **by 15 October 2020 11:00 am (Konstanz time) using the Online Application Platform** https://www.uni-konstanz.de/zukunftskolleg/online-application/.

Details concerning the application, required documents, and information about the Zukunftskolleg are available on our website: https://www.uni-konstanz.de/zukunftskolleg/ .



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

ScienceCareers | AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology

University of Massachusetts Medical School

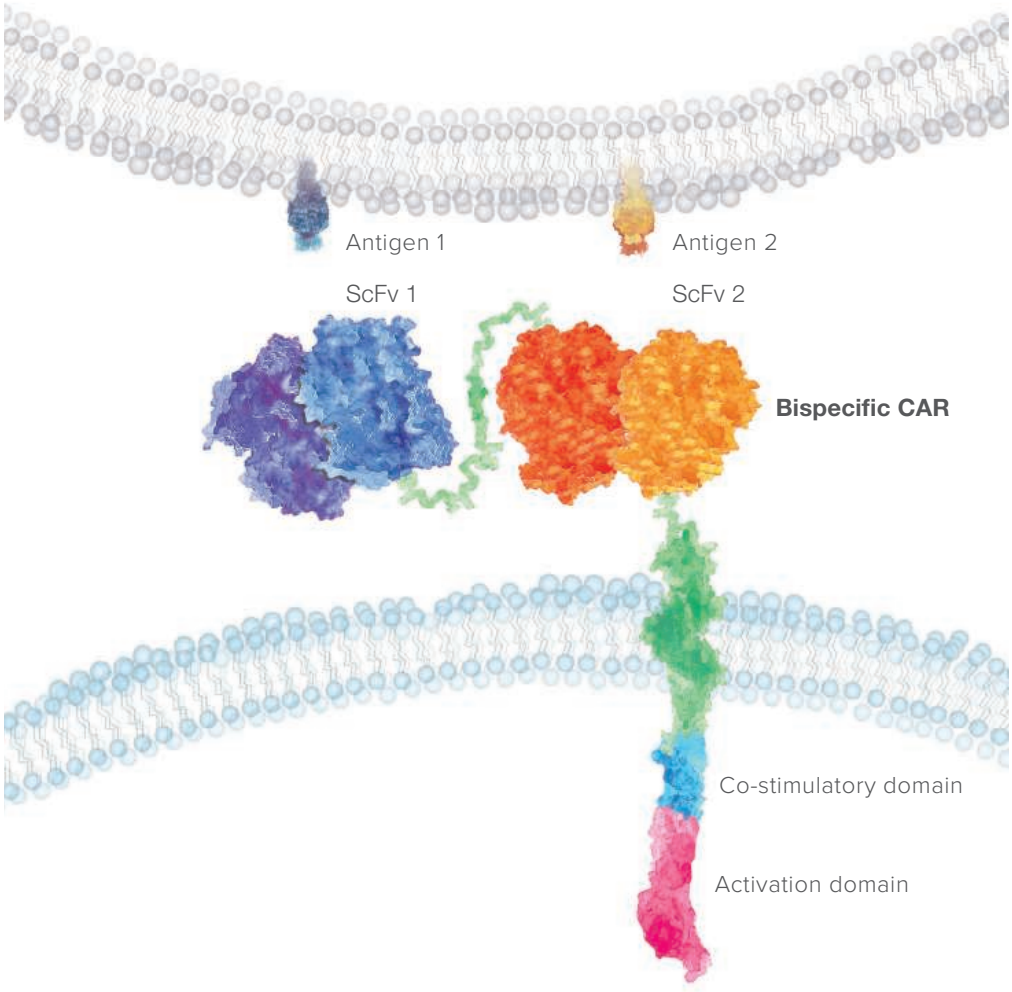
UCSF University of California San Francisco

UIC

BURROUGHS WELLCOME FUND

Custom CAR-T Development

Full Service CRO for CAR-T Research



CAR-T Development Services

- CRO/CMO Services
- Technology Licensing
- IP Development
- Proof of Concept for Clinical Studies
- Preclinical *in vivo* Studies
- Custom Lentivirus Production
- Custom CAR-T Cell Production
- Stable Cell Line Development
- Antibody Development

And many more. Custom services available.

ProMab Biotechnologies allows researchers to utilize our established antibody discovery and CAR-T development platform to streamline their immunotherapy research. With over 100 ready-to-use CAR-T products and several patents in improving CAR technology, we are ready for your next discovery.

All products are for research only

Discover more | www.promab.com

ProMab Biotechnologies, Inc.
2600 Hilltop Dr, Building B, Richmond, CA 94806
1.866.339.0871 | info@promab.com