Does the NIH director have to be an MD? p. 524

Genetic basis for COVID-19 severity pp. 535 & 579

Impacts of Indigenous land dispossession pp. 536 & 578

Science

$15
29 OCTOBER 2021
SPECIAL ISSUE
science.org
AAAS



WHY WE SLEEP



FLEXIBLE.
SENSITIVE.
DEPENDABLE.

CLARIOstar® *Plus*

The CLARIOstar® *Plus* multi-mode microplate reader streamlines assay development and validation by combining monochromator flexibility with best-in-class sensitivity.

- LVF Monochromators™ for highest sensitivity
- EDR technology makes gain adjustment no longer required
- Dedicated detectors for luminescence and red fluorescence
- Best performance in TRF, TR-FRET, FP and AlphaScreen®
- Atmospheric control with gas ramping function
- Made-in-Germany dependability

www.bmglabtech.com

BMG LABTECH

The Microplate Reader Company

©2021 All rights reserved. All logos and trademarks are the property of BMG LABTECH

Science
JOURNALS AAAS



Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**



Chengdu Tianfu International Bio-Town: An emerging world-class city for the medical- and health-related industry

For centuries, Chengdu has been known as "*The Land of Abundance*" and the center of western China. Its rich soil and frequent rainfall enabled agricultural and handicraft industries to flourish there, providing food and tools for the whole country. Today, Chengdu, with a history of more than 2,300 years as a city, has fostered another industry–the medical- and health-related industry–and now its ambition is to benefit the whole world.

Today, Chengdu has 164 million outpatients per year and over 5 million people working in the medical and health-related industry. To fulfill its goal of benefiting humanity, Chengdu has built the Chengdu Tianfu International Bio-Town (CDBT), which is located at the southern part of the city and has an area of 44 km^2 (17 mi^2). From the air, CDBT looks like a garden, featuring lush patches of grass and dense stands of trees. The idea of building the Bio-Town in a forest and building labs in a garden is engraved in the city's genes and stands as one of the many reasons why Chengdu is known as one of most livable cities in China.

Along with the trees found there, projects incubated in CDBT also grow steadily. In the past 3 years, CDBT has authorized about 150 projects, with a total investment volume of over 110 billion CNY (17 billion USD).

Industrial chain as the foundation

Business is the basis of CDBT's development, and CDBT is endeavoring to build a complete industrial chain.

CDBT focuses on developing five major fields–biotechnology drugs, innovative chemical entities,advanced medical equipment, professional outsourcing services, and cutting-edge health care services–and 14 subfields, including antibody drugs, cell and gene therapy, in vitro diagnostics, and smart wearable devices. It has attracted more than 3,200 companies, including AstraZeneca, Medtronic, Johnson & Johnson, Thermo Fisher Scientific, Brilliant Pharmaceuticals, and Maccura Biotechnology. Among them, eight companies, including HitGen and Olymvax Biopharmaceuticals, are listed companies that started in this area, and eight companies, including Westvac, are unicorn companies.

In 2020, CDBT was rated no. 3 in a comprehensive ranking of China's 169 medical- and health-related industrial parks, according to an industry report issued by the China National Center for Biotechnology Development. As the medical and health care industry is facing explosive growth, CDBT plans to realize an industrial scale of 250 billion CNY (39 billion USD) by 2025.

Innovation-driven development: From labs to industrialization

CDBT's innovation-driven development helps boost its growth in many ways. It is China's first Major Pilot Demonstration Base for New Drug Creation and shoulders the national mission of solving the transfer and transformation of innovative medical achievements.

Aware of the importance of working with scientists, CDBT values cooperating with enterprises and institutes in China and around the world to encourage innovation. It has teamed up with Thermo Fisher to establish the Thermo Fisher Scientific Innovation Lab, and with Sichuan University to create the Sichuan University Parallel Lab. Both labs provide enterprises with ready-to-use equipment.

CDBT also innovates development models. It supports principal investigators in transferring their intellectual property to the capital market, aiming at fostering initial public offering companies.

IMAGES PROVIDED BY CHENGDU

To date, CDBT has created 27 platforms for drug screening, discovery, and product manufacturing, and has provided over 1,000 rounds of services for enterprises both at home and abroad.

Upgrading the value chain: Focusing on satisfying clinical needs

Meanwhile, to encourage research that satisfies clinical demands, CDBT has proactively built the Tianfu International Medical Center, which comprises several hospitals and is aimed at building a bridge between labs and hospitals and harboring international multicenter clinical trials. Through the Medical Center, innovative R&D achievements from labs can be transferred to clinical applications in hospitals–achievements that are also based on hospitals' clinical demands in the real world. CDBT has established China's first clinical drug trial investigator association and good clinical practice (GCP) company. It has also established the BOE Digital Hospital and is speeding up construction of the West China Heavy Proton and Ion Hospital as well as the Sichuan Provincial Maternal and Child Health Care Hospital.

Suppy chain: Advantages brought by the "Land of Abundance"

Chengdu has provided CDBT with many advantages, and one of them is preferential location with busy air and ground transportation.

As a critical city in China's *One Belt, One Road Initiative*, Chengdu is now the third Chinese city with two international airports. Since 2017, its Chengdu–Europe Express has been able to reach Lodz, Poland, in 15 days. Chengdu is also the fourth Chinese port approved for biological products import. CDBT is the first industrial zone in China to have a Biomedical Supply Chain Industrialization Service Center. All these factors have made pharmaceutical companies such as Allergan and CSL Behring choose Chengdu as the place for Customs clearance.

Emphasizing the importance of human resources and financial supports

Also critical to the development of CDBT is human resources. CDBT sees highly skilled individuals as the primary core factor in industrial innovation and has cultivated a competitive reward policy for talented researchers. At present, CDBT has gathered more than 60,000 various talents in different industries, including five teams led by Nobel Prize laureates and 10 teams led by academicians.

To make these talents feel at home, CDBT has built a complete production-guaranteed service system in terms of hazardous waste treatment, energy guarantee, and 5G network. It is attentive to the construction of living facilities and is committed to building a *15-minute comprehensive business circle* and a *5-minute basic living circle* (i.e., people can reach their offices in a 15-minute walk and their living facilities, such as apartments, international schools and high-end hospitals, in a 5-minute walk).

CDBT actively creates innovative financial products and models, establishing China's first *biomedical insurance supermarket* and introducing specialized funds. It has also developed a *Bio Loan series* of financial products suitable for early-stage loans, aimed at helping small and medium-sized enterprises expand their financing channels. Meanwhile, the Provincial Medical Products Administration has set up service stations in CDBT that help facilitate faster approval of medical products.

Join us

CDBT is committed to creating value for biological industries worldwide. We sincerely invite experts, entrepreneurs, and everyone interested in promoting the medical and health-related industry to visit and to settle down and participate in its further development.

Sponsored by

BIOTOWN
天府国际生物城



IMAGES PROVIDED BY CHENGDU



You heard the message.

We've told you before that NEB® offers a broad portfolio of reagents for purification, quantitation, detection, synthesis and manipulation of RNA. But did you know that these products are available from bench-scale to commercial-scale to enable both academic and industrial needs? Further, we provide these products at quality levels that support vaccine and diagnostic manufacturing. Experience improved performance and increased yields, enabled by our expertise in enzymology.



RNA purification: Extract up to 100 µg of high quality, total RNA from a variety of sample types with the Monarch® Total RNA Miniprep Kit. Monarch RNA Cleanup Kits can quickly and easily clean up and concentrate RNA in just minutes, with no carryover contamination.



RNA detection: Optimize your RT-qPCR across a variety of sample types with Luna®. High-concentration mixes and kits optimized for multiplexing enable sensitive detection of SARS-CoV-2. Simple, one-step solutions for LAMP and RT-LAMP are also available.



RNA-seq: NEBNext® kits are available for RNA library preparation, rRNA depletion and poly(A) mRNA isolation. Save time with streamlined workflows, reduced hands-on time and automation compatibility.



RNA synthesis: Synthesize high-quality RNA with reagents designed to simplify your workflow, including HiScribe™ IVT kits and capping reagents. GMP-grade* reagents are available for mRNA synthesis of therapeutics and vaccines.

Find more details on products available, request samples, and access helpful RNA-related resources at **www.neb.com/RNA2021**.

*"GMP-grade" is a branding term NEB uses to describe reagents manufactured at our Rowley, MA facility, where we utilize procedures and process controls to manufacture reagents in compliance with ISO 9001 and ISO 13485 quality management system standards. NEB does not manufacture or sell products known as Active Pharmaceutical Ingredients (APIs), nor do we manufacture products in compliance with all of the Current Good Manufacturing Practice regulations.

One or more of these products are covered by patents, trademarks and/or copyrights owned or controlled by New England Biolabs, Inc. For more information, please email us at gbd@neb.com. The use of these products may require you to obtain additional third party intellectual property rights for certain applications.

© Copyright 2021, New England Biolabs, Inc.; all rights reserved.

NEW ENGLAND BioLabs® Inc.

be INSPIRED
drive DISCOVERY
stay GENUINE

CONTENTS

29 OCTOBER 2021 • VOLUME 374 • ISSUE 6567

SPECIAL SECTION

WHY WE SLEEP

INTRODUCTION

550 The many benefits of healthy sleep

NEWS

552 Divided we sleep *By R. Pérez Ortega*

REVIEWS

556 The inescapable drive to sleep: Overlapping mechanisms of sleep and sedation *N. P. Franks and W. Wisden*

560 Brain neural patterns and the memory function of sleep *G. Girardeau and V. Lopes-dos-Santos*

564 The interconnected causes and consequences of sleep in the brain *L. D. Lewis*

568 The transitional neuroscience of sleep: A contextual framework *M. A. Grandner* et al.

ON THE COVER



Sleep is an essential physiological process. Our bodies may be quiet and largely immobile, but our brains are highly active. Metabolites and other toxins that have accumulated over the course of the day get cleared away. Recent memories become consolidated. Scientists have thus begun to investigate the restorative and beneficial functions of sleep in great detail. See the special section beginning on page 550. *Illustration: Dion MBD*

SEE ALSO NEWS STORY p. 526; POLICY FORUM p. 530; VIDEO



ILLUSTRATION: DION MBD

NEWS

IN BRIEF

516 News at a glance

IN DEPTH

518 United Kingdom moves to prevent birth defects
European Union countries are holdouts against widespread practice of adding folic acid to flour *By M. Wadman*

519 Pandemic productivity dip may linger
Scientists started fewer new research projects in 2020 *By J. Madhusoodanan*

520 Superconductor finding draws pointed critique
Study slammed by critic, who has demanded raw data withheld by authors *By R. F. Service*

521 Dinosaurs thrived until asteroid hit, new fossils suggest
Data from Tanis site—claimed to preserve the Cretaceous extinction—impress some geologists, but skeptics remain *By M. Price*

522 Pandemic spotlight brings attention—and plenty of heat
Scientists like Devi Sridhar leaped to fill an information gap *By K. Kupferschmidt*

524 Should the next NIH director have an M.D.?
Some observers say White House should drop that tradition to broaden the candidate pool *By J. Kaiser*

FEATURES

526 The simplest of slumbers
Evidence from evolutionarily ancient creatures is revealing that sleep is not just for the brain *By E. Pennisi*
WHY WE SLEEP SECTION p. 550; PODCAST

INSIGHTS

POLICY FORUM

530 Informing sleep policy through field experiments
Evidence is particularly needed from poorer communities *By G. Rao* et al.
WHY WE SLEEP SECTION p. 550

RECOGNIZE A SCIENTIST WHOSE DISCOVERIES ARE TRANSFORMING MEDICINE

Established by the Feinstein Institutes for Medical Research and *Molecular Medicine,* the **ROSS PRIZE IN MOLECULAR MEDICINE** is awarded annually to an active investigator who has made, and continues to make, an impact in the field through innovative research. Submissions are due November 15, 2021.

The winner will receive a $50,000 honorarium, as well as an invitation to deliver a keynote lecture, at a New York Academy of Sciences symposium in June 2022.

Nominate or learn more at Northwell.edu/RossPrize

Feinstein Institutes for Medical Research
Northwell℠
RAISE HEALTH

PERSPECTIVES

534 Clearing stressed cells
Cell cycle arrest produces a p21-dependent secretome that initiates immunosurveillance of premalignant cells *By V. Reen and J. Gil*
RESEARCH ARTICLE p. 577

535 Defective viral RNA sensing linked to severe COVID-19
Genetic variation in a sensor of double-stranded RNA can exacerbate COVID-19 *By J. Schoggins*
RESEARCH ARTICLE p. 579

536 Documenting indigenous dispossession
A major data synthesis maps tribal displacements *By D. L. Fixico*
RESEARCH ARTICLE p. 578

538 Memories off the top of your head
Cortical layer 1 has a special role in long-term memory
By J. N. Shin et al.

540 Polymer design for the circular economy
Addition of keto groups to polyethylene helps it degrade while maintaining its properties
By M. J. Sobkowicz
REPORT p. 604

BOOKS *ET AL.*

541 The question of canine origins
With few solid leads, a paleoanthropologist explores the emergence of the first dogs
By G. Larson

542 Refining the tree of life
Human interventions, both intentional and unintentional, define a new chapter in Earth's storied history
By S. Campbell-Staton

LETTERS

544 WTO must ban harmful fisheries subsidies
By U. R. Sumaila et al.

544 Alzheimer's drugs: Does reducing amyloid work?
By M. Thambisetty et al.

545 Response
By D. J. Selkoe

RESEARCH

IN BRIEF

574 From *Science* and other journals

RESEARCH ARTICLES

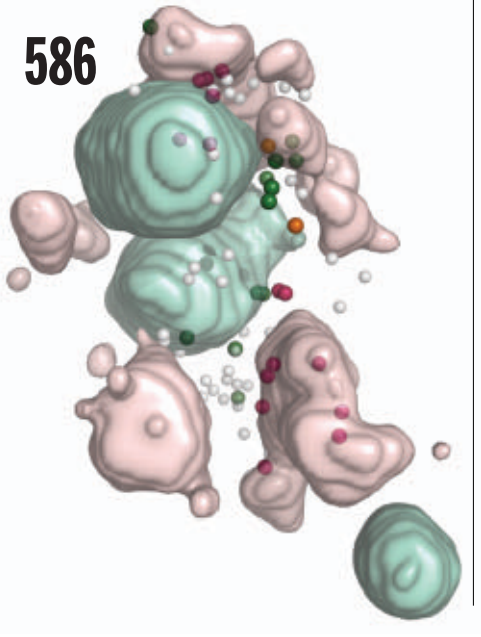
577 Immunology
p21 produces a bioactive secretome that places stressed cells under immunosurveillance *I. Sturmlechner* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABB3420; PERSPECTIVE p. 534

578 Indigenous peoples
Effects of land dispossession and forced migration on Indigenous peoples in North America *J. Farrell* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABE4943; PERSPECTIVE p. 536

579 Coronavirus
A pernylated dsRNA sensor protects against severe COVID-19 *A. Wickenhagen* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABJ3624; PERSPECTIVE p. 535

580 Lipid transport
Distinct allosteric mechanisms of first-generation MsbA inhibitors *F. A. Thélot* et al.

586 Nuclear architecture
Single-cell nuclear architecture across cell types in the mouse brain *Y. Takei* et al.



CREDIT: Y. TAKEI *ET AL.*

REPORTS

594 Phytoplankton
Global drivers of eukaryotic plankton biogeography in the sunlit ocean
G. Sommeria-Klein et al.

599 Sediment transport
Exceptional increases in fluvial sediment fluxes in a warmer and wetter High Mountain Asia *D. Li* et al.

604 Polymer chemistry
Polyethylene materials with in-chain ketones from nonalternating catalytic copolymerization *M. Baur* et al.
PERSPECTIVE p. 540

608 Superconductivity
Direct evidence for Cooper pairing without a spectral gap in a disordered superconductor above T_c
K. M. Bastiaans et al.

612 Plasmonics
Electrically switchable metallic polymer nanoantennas *J. Karst* et al.

616 Magnetism
Intrinsic 2D-XY ferromagnetism in a van der Waals monolayer
A. Bedoya-Pinto et al.

621 Perovskites
Liquid-phase sintering of lead halide perovskites and metal-organic framework glasses *J. Hou* et al.

Plant Science

625 NIN-like protein transcription factors regulate leghemoglobin genes in legume nodules *S. Jiang* et al.

629 Processing of NODULE INCEPTION controls the transition to nitrogen fixation in root nodules
J. Feng et al.

DEPARTMENTS

511 Editorial
Linking climate and biodiversity
By Ramón Pichs Madruga

650 Working Life
Your peers are your pillars
By Katharina Maisel

AAAS News & Notes 549
AAAS Annual Meeting 633
Science Careers 640

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2021 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.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.



Structures of HIV protease (turquoise, PDB ID 3pj6) have been used to design powerful drugs for HIV therapy. Illustrator: Maria Voigt, RCSB PDB

Rutgers Celebrates the 50th Anniversary of the Protein Data Bank

An Essential Global Science Resource Enabling Breakthroughs in Biological and Biomedical Research and Education

Headquartered in the United States at Rutgers University, the Protein Data Bank (PDB) is the primary data resource for structural biology—the world's only free, open-access 3D digital library for proteins, DNA, and RNA: the molecules of life found in all living organisms on the planet. With millions of users from every country and territory worldwide, PDB is a vital public data resource that drives discovery and innovation in subject areas from agriculture to zoology.

PDB Impact and Outcomes

- **180,000+** 3D structures of proteins, DNA, and RNA, and growing every day
- **2 million+** structure data files downloaded daily by users
- **1 million+** published scientific research papers relied on open access to PDB data
- **90%** of FDA-approved anti-cancer drugs were facilitated by PDB structures [2010 to 2018]
- **1,400+** SARS-CoV-2 protein structures in PDB help combat the pandemic
- **$18 billion:** That's the replacement value of data safeguarded by PDB at Rutgers

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

excellence.rutgers.edu

RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank) serves as the U.S. data center for the global PDB archive and makes PDB data available to all data consumers at no charge with no limitations on usage.

RCSB PDB is led by Rutgers, a preeminent national research university and New Jersey's top-ranked public university. Established in 1766, Rutgers is the nation's eighth-oldest higher education institution. Rutgers' RCSB PDB collaborators are University of California–San Diego/San Diego Supercomputer Center and University of California–San Francisco.

EDITORIAL

Linking climate and biodiversity

Climate change and loss of biological diversity are global challenges, linked to each other and to other socioeconomic and environmental challenges. These interlinkages have been discussed by the Intergovernmental Panel on Climate Change (IPCC) and the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), valuable references for multilateral negotiations in the United Nations Framework Convention on Climate Change (UNFCCC), whose Conference of the Parties (COP) convenes in November, and the UN Convention on Biological Diversity (CBD), whose COP convened in October. In the past year, in recognition of common challenges and opportunities, the IPCC and IPBES formally collaborated for the first time. Decision-making processes around these complex issues will be more effective if they integrate responses to climate change, biodiversity loss, and human development gaps at various spatial scales: global, regional, national, and local.

Since 1988, the IPCC has brought together a segment of the international scientific community for periodic assessments that provide rigorous results, drawing on diverse climate change science, which are relevant for policy decision-making. The IPBES has, since 2012, assessed past and foreseeable trends in biological diversity. Through thematic reports and integrated (regional and global) assessments, IPBES has mobilized experts from all corners of the globe on these issues, and has positioned itself in a relatively short time as an institution with an authoritative voice in this matter.

The IPCC assesses the impacts on biodiversity both of climate change and of response strategies to climate change: adaptation and mitigation. The IPBES assessments consider climate change as one of the main direct drivers of biodiversity loss. However, a qualitative leap is required for integrating the knowledge incorporated in those results, through scientific collaboration.

Areas of potential cooperation between the scientific communities linked to the IPCC and IPBES include the development of socioeconomic and environmental scenarios with an integrative perspective—to make visible and quantify, to the extent possible, interlinkages between climate change and biodiversity loss, and their implications for sustainable development. These interlinkages are captured, at least partially, in the 17 UN Sustainable Development Goals. A comprehensive consideration of the Indigenous and local knowledge contributions in responding to these global challenges, as well as the development of research capacities on these topics, especially in developing countries, are other issues to be prioritized as part of the joint efforts.

The UNFCCC and CBD COP events constitute spaces for encounters and confrontation, considering, on the one side, common objectives and, on the other side, various priorities based on prevailing global and regional socioeconomic asymmetries. There is an international consensus that both cases are about global challenges, where we are all part of the problem and must all be part of the solution; but the consensus is broken when it comes to negotiating responsibilities and commitments of each country to guarantee lasting solutions.

> **"The UNFCCC and CBD COP events constitute spaces for encounters and confrontation..."**

Solutions to both problems require immediate actions, but with a long-term perspective and a historical approach. Actions must be equitable and consider multiple socioeconomic and environmental interrelationships evaluated by science. Application of the principle of common but differentiated responsibilities, incorporated into the negotiations in the 1992 Rio Declaration on Environment and Development, is indispensable: "The developed countries acknowledge the responsibility that they bear in the international pursuit of sustainable development in view of the pressures their societies place on the global environment and of the technologies and financial resources they command".

This process of transformative changes would only be viable to the extent that necessary financial resources and technological support are mobilized to recipient regions and countries. As a first step, in the context of the UNFCCC and the Paris Agreement, developed countries should meet and exceed the goal of $100 billion annually in financial support to developing countries, with new pledges for the period 2021–2025. Appropriate balance between climate finance for adaptation and mitigation is crucial for implementing sustainable responses to climate change, with positive impacts in addressing biodiversity loss and other challenges.

–Ramón Pichs Madruga



Ramón Pichs Madruga is director of the World Economy Research Center (CIEM), Havana, Cuba. He is also vice-chair of IPCC Working Group III and a member of the IPBES Task Force on scenarios and models. rpichs@ciem.cu

PHOTO: LIL MARIA PICHS-HERNANDEZ

10.1126/science.abm8739



Turning carbon green

The world is heating up, resulting in unprecedented numbers of "hundred-year" or even "thousand-year" weather events in what seems like a constant stream. The scientific consensus is that if nothing is done to change the pace and direction of greenhouse gas emissions in the immediate future, the damage will be irreparable—if it is not so already.

Each year, 51 billion tons (gigatons, or Gt) of greenhouse gases—considered the main drivers of climate change—are emitted into the atmosphere, of which 37 Gt are CO_2. Those numbers are increasing, with estimates of a possible 50%–100% rise in CO_2 emissions by 2030 if we continue along our current path.

Because of the situation we find ourselves in today, it is not enough just to reduce emissions and maximize the implementation of renewable sources of energy, which many nations are aiming to do by 2050. There is just too much CO_2 already in the air. "We really need to target negative emissions, meaning that net-zero emission is necessary, but it's not enough, and we need to complement that with a bounty on capturing CO_2 from the atmosphere," says Veronica Bermudez, head of the Energy Center at Qatar Environment and Energy Research Institute (QEERI), part of Hamad Bin Khalifa University (HBKU) in Doha.

Upcycling

QEERI's Carbon Capture and Utilization (CCU) "cluster project" was created and funded by a 5-year, USD 5 million grant from the Qatar National Research Fund (plus

"We really need to target negative emissions...but it's not enough, and we need to complement that with a bounty on capturing CO_2 from the atmosphere."

–Veronica Bermudez



PHOTOS: PROVIDED BY QATAR FOUNDATION

$1.4 million industry cofunding)—at least 80% of which is to be spent in Qatar. The goal is to bring together experts in science, engineering, policy, trade, transportation, and economics from Qatar and around the globe to find ways to feasibly upcycle CO_2 and mitigate its burden (upcycling means transforming something into a more valuable material than the original). The cluster currently boasts 22 local and 11 international investigators drawn from various research institutions and commercial partners, notes Bermudez, who is involved with its technical assessment and management.

Creating an economic advantage

CO_2 can be a valuable commodity in and of itself—to sell to the beverage and agricultural industries, for example—as well as for feedstock to produce other commodities. Yet less than 1% of anthropogenic CO_2 is utilized by industry. Part of the reason for this is that "to date, capturing CO_2 is really expensive," and the process cannot compete today in the real market, Bermudez says.

Bermudez does not believe that "things will happen just because they are the right thing to do; we need to have some economic drivers and a business model behind it," she explains. Part of the answer is to incentivize sustainable strategies so they become economically viable.

Fortunately, the market itself is changing. "With the upcoming carbon border tax, all the products that we plan to import into our countries will be taxed depending on their carbon footprint," Bermudez points out, noting that "Japan has already instituted such a tax, Europe is on the way, and the United States will be doing so very soon." Thus, the cost of releasing CO_2 into the atmosphere will need to be figured into the overall cost of production.

One solution is to turn captured carbon directly into value-added products instead of storing it—a principal goal behind QEERI's CCU project. Treating carbon-

BLUE HYDROGEN

"Circular economy" isn't a new term, but it does seem to be heard much more often these days, and for good reason. "Blue hydrogen energy"—energy produced by a partially renewable process, and with a lower environmental footprint—isn't new either. Taking industrial waste and using it to produce something of value is a goal to strive for.

These are the bases of a collaboration between researchers at Qatar Shell Research and Technology Centre and Konstantinos Kakosimos, associate professor of chemical engineering at Texas A&M University at Qatar. Supported by a grant from the Qatar National Research Fund (QNRF), they are investigating using solar energy to photochemically split wastewater, liberating hydrogen that can be used for fuel. At the same time, they are making use of one of the industrial byproducts of converting natural gas to fuels and other chemicals, says Kakosimos. Typically, water-splitting efforts involve pure water, which is a valuable resource. By contrast, "We're taking what was a waste stream that costs money to treat and using it as a raw material to produce something of added value," explains Kakosimos.

He sees this work as a compromise between the academic drive for perfection regardless of the practicalities of implementation, and the reality of industry's need for feasibility, which takes into account cost and environmental footprint, for example.

Qatar's national goal of sustainability is more attainable because in Doha, all the critical intellectual components—academic research, R&D, technology development, and industrial partners—are so close to each other. "It's like an ecosystem" that can foster both scholarly work and industrial collaboration, Kakosimos points out.



PHOTO: PROVIDED BY QATAR FOUNDATION

containing waste using processes such as mineral carbonation—sequestering CO_2 into minerals such as calcium carbonate, which is used in the building industry—will have the dual benefits of reducing waste disposal and generating useful products that can be reinjected into a sustainable economic cycle.

Split and catch

Air conditioners (ACs) are an off-the-shelf technology that has been around for years. Rather than being constructed as a single unit, many of the latest models are split into two basic parts—the first an internal unit that picks up heat from an enclosed space, and the second an outdoor unit that ejects heat to the outside—with refrigerant flowing between them.

It's possible for ACs to be retrofitted with technology based on CO_2 "nanoscavengers," says Nashaat Nassar, professor of chemical and petroleum engineering at the University of Calgary. One CCU program is developing new, cheaper, and higher-efficiency membranes that can be integrated into AC splitters—and ultimately industrial heating, ventilation, and AC (HVAC) systems—to scavenge CO_2. The captured CO_2 can then be mineralized directly into raw materials such as calcium carbonate and magnesium carbonate, for example. The making of cement is currently a very high CO_2-emitting process, and utilization of captured CO_2 could offset much of those emissions.

An AC will have a filter (or series of filters) through which incoming air will pass. "The first thing you want to do is

GRAY MATTER

Qatar doesn't just invite scientific talent to collaborate or relocate. The petroleum-rich Gulf nation is also tapping—and creating—some of the region's best homegrown minds.

"As the national funding agency, the Qatar National Research Fund (QNRF) invests heavily in building much-needed local human research capacity across all fields of science and all education levels, from K–12 to established scientists," says Abdul Sattar, founding director of QNRF.

Qatar now boasts two homegrown research universities—Hamad Bin Khalifa University (HBKU) and Qatar University—as well as branch campuses of six international universities, all in Doha's Education City.

The national vision

The country's burgeoning research and educational programs tackle some of the world's most pressing issues. Qatar's National Vision 2030 calls for taking a proactive and significant role in assessing and mitigating the impact of climate change. There is an emphasis on using "clean" modes of transportation and renewable sources of energy as much as possible.

As part of his students' graduate training, Yusuf Bicer, assistant professor of sustainable development at HBKU, has them investigating ways to convert solar energy into more storable, transportable, carbon-free forms of energy, such as clean hydrogen and ammonia.

Renewables tend to suffer from intermittency. Bicer says, "We need to utilize and store wind energy whenever there is an excess wind, and do the same with solar energy whenever we have daylight."

Too much sun?

Qatar and the Gulf region have an excess of solar irradiation. This creates both opportunities and challenges for clean energy.

Qatar is a major producer of liquid natural gas (LNG), the combustion of which releases CO_2 and contributes to climate change. Bicer's lab is looking into using solar energy to turn LNG into "a cleaner type of medium: carbon-free hydrogen or ammonia," he says. These can be used for zero-carbon energy storage and fuel. "Transporting electricity overseas using wires exceeding a certain distance is not possible, but we can transport chemicals as energy carriers."

All it's cracked up to be

"The hydrogen economy seems to be the core of future sustainable energy systems," says Amro Mohamed, a chemical engineer and Ph.D. candidate under Bicer.

Producing so-called "dirty hydrogen" uses coal or other nonrenewable energy sources. "My main thesis work is developing and designing materials for photocatalytic reduction of nitrogen," utilizing photons from solar light to break chemical bonds and directly produce green ammonia, Mohamed explains.

He uses rational design, simulations, and experiments to select materials with high thermal stability, relative ease of operation, and most importantly, photosensitivity. "We also need to look at environmental impacts, costs, and logistics," Mohamed says.

(Molten) salt and batteries

Lithium-ion batteries are the most popular way to store energy. But these degrade with temperature over about 40°C, and ambient temperatures in the Gulf region can regularly top 50°.

Manal AlShafi, another graduate student working with Bicer, compared a large variety of mechanical, thermal, chemical, and electromagnetic energy storage methods in terms of a variety of factors, and asked an international panel of experts to weight these factors. We then "compared the environmental impact of storing one kilowatt-hour of energy," she says.

"Basically, we are trying to develop new parameters specifically for hot climates and choosing which energy storage method is more preferable for which renewable energy source by considering these criteria," Bicer says. One may be more efficient, but that can be counterbalanced by its impact on the climate.



“The scale of the climate-change challenge is vast, and this work is a potential game changer to make carbon capture more scalable and economically viable.”

–Marc Vermeersch

remove the big things, like dust, using a physical filter,” says Stefano Sanvito, professor of condensed matter theory at Trinity College Dublin, who has worked with Qatar on energy-related projects for the past 10 years.

But you’re still left with a mixture of gases. “My particular goal is to develop the material that sequesters the CO_2,” he says. “This is tricky, because the molecules are an order of magnitude smaller than a dust particle, so we have to design a new compound.” He and his team are investigating a class of materials called metal organic frameworks, which include a transition metal ion and can be assembled like a Lego to make a highly porous material “which looks and acts like a microscopic sponge.”

They are building on a database of about 40,000 known metal organic structures, none of which are completely suitable for one reason or another. “The idea is to use this vast dataset to design a new or modified metal organic framework that has larger uptake and is more selective for CO_2,” Stefano explains.

“You have to be very careful, in the sense that this air is ultimately going to go to the lungs. So the material you’re choosing for this purification has to be environmentally friendly, nontoxic, and stable,” points out Nassar.

Integration with existing technology

If the amount of CO_2 is appreciable, then it can be recovered as pure CO_2 after regenerating the filter (which can be done by heating) and stored in a specialized receptacle called an ISO (International Organization for Standardization) tank. “It’s a well-known technology,” says Nassar, one of the principal investigators in the cluster. “This can be taken for a process where we can convert CO_2 into a value-added product. And again, the conversion of CO_2 to value-added product is now an off-the-shelf technology. It’s not a new thing—the integration of the processes is the new thing.”

In terms of CO_2 capture and recovery, Qatar benefits from both its small size and the concentration of most of its populace into compounds, “meaning that they are centralized and that we can have quick access for this CO_2 to be transported to the transformation unit,” Bermudez points out. “You don’t have to have a pipeline,” adds Nassar. Large buildings, malls, and universities with their own air intake can easily have an ISO integrated into their HVAC systems to temporarily capture and store CO_2.

The idea is to take advantage of existing infrastructure. Otherwise, it’s really difficult to get something adopted, says Bermudez. “So we are developing new materials that aim to be cheaper and highly efficient, which will be integrated within the AC splitters and HVACs.”

“The scale of the climate-change challenge is vast, and this work is a potential game changer to make carbon capture more scalable and economically viable,” says Marc Vermeersch, QEERI’s executive director. “Qatar has the potential to be a leader in carbon capture technologies embedded in air-conditioning systems, which is another reason this project is so important for QEERI.”

Sponsored by

مؤسسة قطر
Qatar Foundation
Unlocking human potential لإطلاق قدرات الإنسان

PHOTOS: PROVIDED BY QATAR FOUNDATION

NEWS

> “It’s highly unlikely that it is really from a transmitter out at Proxima Centauri.”
>
> **Andrew Siemion of the Breakthrough Listen project,** on a striking radio signal that seemed to come from a nearby star but now appears to be human-generated interference.



IN BRIEF

Edited by **Jeffrey Brainard**

GLOBAL WARMING

Climate summit opens as carbon dioxide cuts lag

As world leaders converge on Glasgow, U.K., for a climate summit beginning 31 October, a series of reports show the world is far from meeting a promise to try to hold global temperature increases below the 1.5°C target that could avert the worst impacts of climate change. To meet the temperature target, set in the 2015 Paris climate agreement, nations need to cut greenhouse gas emissions by 55% by 2030, but current pledges would trim emissions by only 7.5%, the United Nations Environment Programme (UNEP) warned this week. Meanwhile, countries’ detailed plans show they expect to increase fossil fuel production until at least 2040, despite commitments to reduce emissions to net-zero by 2050, a coalition of research institutes and UNEP reported on 20 October. “Countries have not yet done the work of translating what their climate pledges mean to [management of] fossil fuel markets,” says co-author Peter Erickson of the Stockholm Environment Institute. At the Glasgow meeting, countries are expected to increase their targets for emissions cuts for the first time since the Paris agreement established initial targets and a schedule for tightening the commitments.

En route to the climate summit in Glasgow, U.K., Bamber Hawes (right) walks with his sculpture *Clarion the Polar Bear* to raise awareness of the climate crisis.

PHOTO: STUART WALKER/I-IMAGES/EYEVINE/REDUX

More vaccine boosters authorized

COVID-19 | The U.S. Food and Drug Administration (FDA) last week opened the floodgates for people in the United States to receive booster shots of COVID-19 vaccines. Building on a September decision that allowed widespread use of third shots for the Pfizer vaccine, FDA now says data support a third shot of Moderna’s similar product, at least 6 months after the first two. Both the Pfizer and Moderna vaccines contain messenger RNA that codes for the viral surface protein called spike. Everyone age 65 and older is eligible, along with people as young as 18 if they are at high risk of severe COVID-19, are exposed at work, or live in nursing homes, prisons, or other institutions. FDA also decided that everyone 18 and older who received Johnson & Johnson’s vaccine, which uses an adenovirus to deliver a gene for spike and was authorized as a single dose, can now receive a second shot. And FDA gave mixing and matching a green light: People who have completed the primary course of injections and are eligible for boosters can receive any of the three vaccines to fortify their immune responses.

Sensors spot Mexico monuments

ARCHAEOLOGY | A large airborne laser mapping survey of Maya and Olmec sites in southern Mexico has revealed nearly 500 Mesoamerican monuments. Several of

the newly identified ceremonial platforms were so large that ground-based researchers had missed that they were humanmade. Researchers used publicly available lidar data to map structures dating back to 1050 B.C.E. across an 84,500-square-kilometer region. They found evidence for a standardized layout for the platforms based on Mesoamerican cosmology. The design, which may have first appeared at the ancient Olmec site in San Lorenzo, exerted an architectural and cultural influence on later Maya sites across the region, the team reports this week in *Nature Human Behaviour*. The team also found four additional layout types, representing either different cultural influences or different points in time.

MIT biologist sues for defamation

#METOO | David Sabatini, the prominent biologist who resigned from the Whitehead Institute for Biomedical Research in August after a sexual harassment investigation, last week sued the institute and its director for defamation. The lawsuit, filed in Massachusetts state court, also names a junior colleague who complained that he sexually harassed her. In a 72-page complaint, Sabatini alleges that he ended a consensual affair with the colleague, who ran a different Whitehead lab, and that she then "became determined to destroy" him and fabricated the harassment allegations. "The Whitehead's investigation was a sham," it says. Sabatini was also fired by the Howard Hughes Medical Institute in August and placed on administrative leave by the Massachusetts Institute of Technology, which may revoke his tenure. He has lost positions at multiple biotechnology firms, according to the lawsuit, which seeks unspecified damages. The Whitehead Institute and its director, Ruth Lehmann, declined to comment.

Satellite will eye birth of elements

ASTROPHYSICS | NASA has selected a gamma ray telescope as the latest small astrophysics mission in its long-running Explorers program, the agency said last week. Scientists will use the $145 million Compton Spectrometer and Imager, to be launched in 2025, to study how stars forge elements, by tracking gamma rays from the decay of radioactive atoms produced in supernovae. The mission will also study other high-energy objects, such as giant stars exploding as supernovae and matter-gorging black holes in galactic centers. And it will probe the mysterious origin of positrons—the antimatter counterparts of electrons—seen close to the Milky Way's center. The project team, based at the University of California, Berkeley, spent decades developing the telescope by flying prototypes on high-altitude balloons.



Ancient DNA from this woman and other mummies found in northwestern China indicates they were not immigrants, as was thought.

PHOTO: WENYING LI/XINJIANG INSTITUTE OF CULTURAL RELICS AND ARCHAEOLOGY

ARCHAEOLOGY

China's mysterious 'Western' mummies were locals

With their red, blond, and brown hair and felt hats, the renowned Tarim Basin mummies, first discovered in the early 20th century, have long been seen as Westerners who migrated to a desert oasis in northwestern China 4000 years ago. Now, their ancient DNA reveals they were local people with deep roots in Asia, the last known members of an ice age population that was once widespread across Eurasia. But although they were inbred and genetically unrelated to neighboring populations who migrated to the region, they nonetheless were culturally cosmopolitan by the second millennium B.C.E.—adopting woolen clothing, wheat, barley, sheep, goats, and cattle from unrelated herders, according to a study in this week's issue of *Nature*. An international team examined the genomes of 13 of the oldest mummies, all of which had been naturally mummified by the basin's dry, salty environment.

Predatory publisher rebrands

PUBLISHING | OMICS, a company based in India that U.S. authorities hit with a $50 million judgment in 2019 for deceptive and predatory publishing practices, has rebranded its journals to avoid the stigma, a study in this week's issue of *Nature* reports. Most of its more than 700 journals are now published by OMICS subsidiaries, and some articles were backdated to before the new subsidiaries were created. Among the practices that drew the U.S. judgment was charging authors fees for publishing articles open access without providing peer review; the study's authors suggest "starving the Hydra" of these funds by pressing OMICS and other predatory publishers to post their peer reviews online.



Melanie Chambers (right) and her daughter Ella, 17, who has spina bifida. They helped push the United Kingdom to fortify flour with folic acid.

PHOTO: PETER SANDGROUND/SBH SCOTLAND

PUBLIC HEALTH

United Kingdom moves to prevent birth defects

European Union countries are holdouts against widespread practice of adding folic acid to flour

By **Meredith Wadman**

In 2013 at the Scottish Parliament in Edinburgh, 9-year-old Ella Chambers of Glasgow spoke to politicians from her wheelchair. Ella has spina bifida, a birth defect that occurs when the embryonic neural tube does not close properly during the first month of pregnancy. She and her family were campaigning for the U.K. government to mandate fortification of flour with folic acid, a synthetic form of the B vitamin folate that can prevent many such defects. Ella's mother had taken folic acid supplements—but not until she knew she was pregnant, by which point the damage was done.

The United Kingdom studied and wrangled over the issue for another 8 years. Finally, last month, the country announced it would begin the mandatory addition of folic acid to white flour. New Zealand made a similar announcement in July. The U.K. decision, the culmination of a long campaign that included pro-fortification recommendations by government science advisers in 2006 and again in 2017, puts the country in line with 80 or so others, including the United States, Canada, and Australia, many of which made the change decades ago.

"I couldn't believe they had actually done it," says Ella, now 17, of the U.K. move. "It was a bit of a shock—in a good way."

"There's nothing that can help Ella," who is paralyzed from her chest down and has learning difficulties, says her mother, Melanie Chambers. The family has worked with Spina Bifida Hydrocephalus Scotland to promote fortification. "But other families hopefully won't have to go through what we have."

Yet the nations of the European Union are not following suit. For reasons that include a wariness about unnatural ingredients in food, they have steadfastly declined to mandate flour fortification. The average prevalence of neural tube defects (NTDs) in 19 European countries is about 10 for every 10,000 pregnancies, notably higher than in others where fortification is required. (The figure is seven per 10,000 pregnancies in the United States [see graph, p. 519].)

"Europe is really dragging the chain," says Lyall Thurston, a veteran New Zealand activist whose 38-year-old son has spina bifida. "It's incredibly sad. Because the science and the research is there."

Babies with NTDs are born without brains or with their spinal cords protruding from their backs, the former killing them and the latter often causing lifelong disability. Many pregnancies are terminated when an NTD is diagnosed. An estimated 70% to 80% of cases are due to inadequate folate, an essential nutrient that promotes DNA synthesis and is present in leafy greens, liver, lentils, and other dietary sources.

Back in 1991, Nicholas Wald, now a professor of preventive medicine at University College London, and colleagues showed that folic acid supplements taken before and during early pregnancy substantially reduced NTDs. Folic acid is a synthetic form of folate that is better absorbed than the natural form and more effective in preventing NTDs. Because the neural tube closes by 28 days after conception, before most women know they are pregnant, starting in the 1990s many countries aimed to boost the nutrient value for everyone. They required companies to add folic acid to white flour. (In the United States, "enriched" rice, pasta, and cornmeal are also fortified.) The numbers of pregnancies affected by NTDs dropped in every country studied.

But no EU nation joined them. Fortification "has not been considered justified" by the scientific evidence, says an official of the European Commission. Even targeted supplementation isn't uniformly urged by European policies, most of which date to the 1990s. In a survey published in 2016, only three of 20 countries recommended that all women of childbearing age take a daily folic acid pill. Although 11 countries recommended folic acid supplementation by women planning a pregnancy, none specified that they should start 12 weeks before conceiving, the time needed to attain optimal protective levels, according to work published since 2000.

Such measures are inadequate in any case, fortification advocates say. They note that about 50% of pregnancies are unplanned and many women, particularly disadvantaged ones, don't know they need supplements.

The result is an elevated rate of NTDs in much of Europe—in Bulgaria, for example, 22 of every 10,000 pregnancies are affected. The rate is lower in a few countries such as

Italy (5.4), perhaps because of diet or widespread use of folic acid supplements by women before becoming pregnant.

Pressure to fortify in Europe may have been muted because most affected fetuses are identified by antenatal screening and terminated, health experts say. Across 19 EU countries that are full members of a European Commission registry, 80% of affected pregnancies were terminated in recent years. (In the United States, an estimated 23% to 42% are terminated.)

Compared with many other countries, "Europe has very good access to terminations," says Vijaya Kancherla, an epidemiologist at the Center for Spina Bifida Prevention at Emory University. "There are not as many babies born with these conditions, so they don't see them and they think there is no problem."

Kancherla adds that many Europeans "see anything added to food as … not natural, as being dangerous to the body. Which is not true for folic acid."

Some European experts are wary of potential adverse effects from excessive folic acid intake in the general population. That's what the German Federal Institute for Risk Assessment (BfR) wrote in a 2017 report rejecting mandatory fortification. BfR recommended folic acid intake be boosted only in the target group of women. "The benefits of folic acid for a relatively small group must be weighed against the possible disadvantages for a large part of the population," the report argued.

Adverse effects have been seen at very high doses of folic acid, in excess of 5 milligrams daily, or 36 times the amount added to a 100-gram serving of fortified white bread in the United States. In the 1940s, some people with vitamin B12 deficiency given very high doses of folic acid to treat anemia developed irreversible neuropathies. More recent studies have hinted at cognitive impairment in the elderly and growth promotion of existing cancers. But other studies have found that folic acid protects against some cancers. "Strong biological and mechanistic premises connecting [high folic acid levels] to adverse health outcomes are lacking," a 2019 National Institutes of Health workshop concluded.

"We have had folic acid fortification in the U.S. for [23] years and we haven't seen skyrocketing rates of colon cancer or other conditions," adds Philip Lupo, an epidemiologist at Baylor College of Medicine.

Irwin Rosenberg, a physician and nutritionist who is a professor emeritus at Tufts University, says he's worried about excess folic acid from supplements in the United States, where supplements are popular and largely unregulated. But he's not concerned about fortified flour. The potential adverse effects of fortification are so weak that they "really should not have been used … as an argument against fortification," he says.

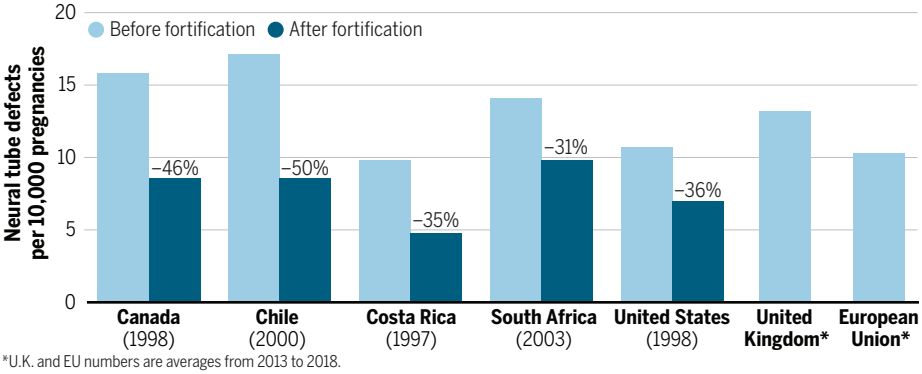
"The evidence against folic acid [fortification] is really weak," agrees Rima Obeid, a clinical biochemist at Saarland University Hospital in Homburg, Germany. "The problem is you don't have [a] political interest" in getting it done.

Possible toxicity from folic acid is the last thing on the mind of Katya Kovacheva, of the Medical University, Pleven, in Bulgaria, who runs the country's only birth defects registry. "I have drawn attention to the high prevalence of NTDs in our country … unfortunately without a response from the authorities," she says.

Michael Turner, a professor of obstetrics and gynecology at University College Dublin, says lobbying authorities country by country is unlikely to yield results. "I would make a heartfelt cry for all European countries … to come together to standardize mandatory food fortification in Europe. There needs to be a pan-European solution to this." ■

The power of folic acid

The prevalence of neural tube defects dropped notably in the years after these countries began to fortify flour with folic acid. The European Union does not require fortification and the United Kingdom just decided to do so.



*U.K. and EU numbers are averages from 2013 to 2018.

CREDITS: (GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) EUROPEAN COMMISSION; NATIONAL CENTER ON BIRTH DEFECTS AND DEVELOPMENTAL DISABILITIES/U.S. CENTERS FOR DISEASE CONTROL AND PREVENTION

COVID-19

Pandemic productivity dip may linger

Scientists started fewer new research projects in 2020

By **Jyoti Madhusoodanan**

The pandemic led to a surge in COVID-19 research, but it severely disrupted other fields—shuttering labs, restricting travel, and leaving scientists with children struggling to work without adequate child care. A flurry of studies indicates the productivity of women scientists slowed during the pandemic by more than that of their male colleagues. Now, a survey of a large swath of the scientific community reveals those impacts may be felt for years to come because many investigators—especially women and scientists with young children—were unable to start new research projects in 2020.

"There may be a cliff looming on the horizon," says Dashun Wang, a computational social scientist at Northwestern University who headed the new study, published this week in *Nature Communications*.

Wang's team polled 7000 U.S. and Europe-based principal investigators in January, asking how their research activity in 2020 compared with 2019. Respondents who did not study COVID-19 reported a drop in papers published (–9%) and manuscripts submitted (–15%). But the most glaring falloff came earlier in the research pipeline, as there was a 36% decrease in the initiation of projects. Women and parents of children 5 years of age or younger were especially hard hit.

The findings are crucial to informing policies designed to support the careers of researchers impacted by pandemic disruptions, says Catherine Wagner, an evolutionary biologist at the University of Wyoming who has written about the challenges she and other scientist mothers have faced. In addition to offering tenure clock extensions, for example, institutions could better support child care and offer seed grants that support new collaborations, she says. "It's really important to consider how COVID-19 will impact productivity far into the future." ■

Jyoti Madhusoodanan is a science journalist in Portland, Oregon.

HIGH-PRESSURE PHYSICS

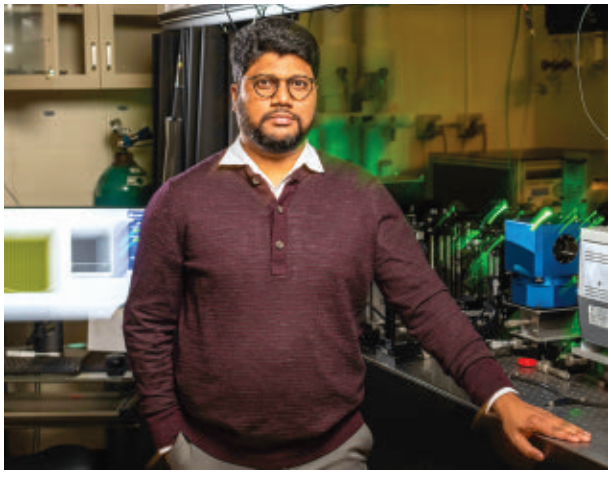
Superconductor finding draws pointed critique

Study slammed by critic, who has demanded raw data withheld by authors

By **Robert F. Service**

A result last year hailed as a breakthrough in physics also generated skepticism that has now escalated into angry recriminations. Researchers said they had made the first superconductor that works at room temperature, a long-sought goal. But Jorge Hirsch, a physicist at the University of California (UC), San Diego, attacked some of the evidence, particularly a set of magnetic measurements. He says his requests to see the underlying data have been rebuffed by the authors for nearly a year. And, last month in a peer-reviewed paper, he charged that the results are "probably fraudulent."

Ranga Dias, an applied physicist at the University of Rochester, who with his colleagues made the room-temperature superconductivity claim, rejects Hirsch's allegations. He asserts that Hirsch isn't an expert in high-pressure physics and that he has a history of claiming that the Nobel Prize–winning "BCS theory" underlying superconductivity is incorrect. Dias says Hirsch relentlessly badgers superconductivity researchers. "Hirsch is a troll," Dias says. "We are not going to feed this troll" by providing the data.



Ranga Dias's room-temperature superconductor paper is under fire.

Superconductivity is normally seen only at temperatures well below 200 K, or −73°C. But several research groups working with hydrogen-rich compounds called hydrides have reported that they became superconductors between 200 K and 250 K when squeezed to intense pressures. Dias and his team went further. They reported that by adding a bit of carbon to precursors of H_3S, a known hydride superconductor, they were able to create a carbon sulfur hydride (CSH) material that pushed the superconducting temperature up to 287 K (14°C), the temperature of a cool room. The result, published in the 15 October 2020 issue of *Nature*, drew worldwide acclaim (*Science*, 16 October 2020, p. 273).

Some scientists attempted to replicate or extend the finding, without much success. And Hirsch and others raised concerns (*Science*, 27 August, p. 954). Like other superconductors, CSH showed a characteristic plunge in electrical resistance as it dropped below the "critical temperature" (T_c) and became a superconductor. To confirm that a material superconducts, however, physicists also look for a second telltale indicator, known as the Meissner effect, in which the material expels magnetic fields below T_c.

Measuring the Meissner effect hasn't been possible in hydrides because they are formed in minute amounts inside a high-pressure device called a diamond anvil cell, which normally is made from magnetic materials. So, hydride researchers have instead evaluated a property known as AC susceptibility, a measure of how much a material becomes magnetized in an applied magnetic field. In Dias's *Nature* paper, CSH's AC susceptibility dropped sharply at T_c, consistent with the interpretation that the material was expelling magnetic fields.

But the data also show that as the material cools below T_c, the AC susceptibility rises again. That's a behavior not usually seen in superconductors, Hirsch argues, though others say the behavior has been seen in other superconductors under high pressure.

Hirsch also contends some of the AC susceptibility data for CSH look suspiciously similar to other data that are now in question, from a 2009 *Physical Review Letters* paper on superconductivity in europium under high pressure. One of that study's authors is James Hamlin, a physicist now at the University of Florida, who participated in the europium study as a graduate student. He recently determined that "there are alterations to the data." The study's senior author, James Schilling, an emeritus physicist at Washington University in St. Louis, says he shares that concern. Now, some of the coauthors are redoing the measurements, and if they don't hold up, the team will retract the paper, Hamlin says.

The first author of the europium paper, Mathew Debessai, now with Intel Corporation, was responsible for the AC susceptibility measurements, and also carried out those measurements for Dias's CSH work. And a data trace from the CSH paper looks "remarkably similar" to one from the europium paper, Hirsch contends. Schilling agrees parts of the traces do look similar, but he can't say why. Debessai declined to respond to questions about the data but via email wrote that he will post a formal response on the arXiv preprint server.

In October and November 2020, Hirsch emailed Dias requests for raw data. Dias replied that he would not provide the data for reasons including that his team was working to patent the work, and his lawyer advised against releasing the data. By then, Hirsch had raised concerns about the CSH data in a preprint, which was published in *Nature* in August. In his email to Hirsch, Dias wrote, "Given that you have an active comment on our work, we consider such a request would not be reasonable."

Frustrated, Hirsch requested the data from *Nature* and the National Science Foundation (NSF), which funded the work. On 30 August, *Nature* appended a note to Dias's paper saying: "The editors of *Nature* have been alerted to undeclared access restrictions relating to the data behind this paper. We are working with the authors to correct the data availability statement." NSF and the University of Rochester both tell *Science* they cannot comment on possible investigative matters.

Then, last month in *Physica C: Superconductivity and its Applications*, Hirsch wrote that Dias's anomalous AC susceptibility result, combined with seeming irregularities in the europium data, and his struggles to get the data on CSH, fed his doubts. "I suggest that one possible explanation is that [the CSH finding] is the result of data manipulation and alteration such as was described … for [europium]," Hirsch wrote.

Alexander Goncharov, a physicist at the

PHOTO: ADAM FENSTER/UNIVERSITY OF ROCHESTER

Carnegie Institution for Science, won't go that far, but he thinks Hirsch's concerns about AC susceptibility measurements are fair. "I tend to believe there is a problem with the [CSH] paper," says Goncharov, who tried repeatedly to synthesize CSH using Dias's recipe and failed.

Given the questions, several scientists say Dias should make his data public. "I am unhappy that Dias is supposedly not cooperating with researchers who are questioning his data," says Marvin Cohen, a theoretical physicist at UC Berkeley. Schilling is blunt: "I told Dias to give [Hirsch] the raw data, for heaven's sake."

Dias and others say they don't trust Hirsch to appraise the data fairly. "Unfortunately, sometimes he is not objective," says Vasily Minkov, a chemist at the Max Planck Institute for Chemistry who synthesizes hydride superconductors and says Hirsch has cherry-picked data from the Max Planck experiments for his critiques. Hirsch calls such critiques "completely unfounded." He says his belief that the BCS theory is incorrect "does not mean I am 'biased' or not 'impartial.' It means that I am motivated to scrutinize carefully the experimental evidence and judge it on its merits, as opposed to assuming it is likely to be right because BCS theory predicts it to be right, as everybody else does."

Several new results have only deepened the mystery. In a preprint posted on 30 September on arXiv, Goncharov's group reports synthesizing CSH under high pressure—using a recipe different from Dias's—and observing a crystalline structure similar to the one Dias reported. Goncharov's team didn't test whether its CSH sample was superconducting. But in unpublished work, Minkov says he and his colleagues synthesized the same CSH structure as Dias's and found it doesn't superconduct above the T_c of H_3S. Minkov says H_3S may be responsible for superconductivity in his CSH sample. "We couldn't see any effect of carbon," he says.

Hirsch, meanwhile, is mounting a broader attack: on claims for superconductivity in any hydride. In a preprint posted on 4 October on Research Square, Minkov and a team led by Mikhail Eremets, a physicist also at Max Planck, reported remaking diamond anvil cells without magnetic materials and testing large samples of two hydrides, H_3S and LaH_{10}. The result: the first ever evidence of the Meissner effect in hydrides, which the team calls "unambiguous evidence" that superconductivity in hydrides is real. Hirsch disagrees, calling the analysis "deeply flawed" in a preprint he and a colleague posted on 14 October on the arXiv server.

Only one thing seems certain to emerge from the controversy over room-temperature superconductivity: more heat. ■



PHOTO: ROBERT DEPALMA

Robert DePalma (left) and a colleague study a dinosaur footprint, likely from a theropod, at North Dakota's Tanis site.

PALEONTOLOGY

Dinosaurs thrived until asteroid hit, new fossils suggest

Data from Tanis site—claimed to preserve the Cretaceous extinction—impress some geologists, but skeptics remain

By **Michael Price**

When a massive asteroid struck Earth some 66 million years ago, were dinosaurs around to experience the cataclysm? Two years ago, a paleontologist claimed to have found evidence at a fossil-rich North Dakota site called Tanis that dinosaurs were alive until moments after the impact, when floodwaters surged over them. But many paleontologists were skeptical, especially because the dinosaur data were first discussed in a magazine story rather than a peer-reviewed journal.

Last week, at the annual meeting of the Geological Society of America in Portland, Oregon, paleontologist Robert DePalma and colleagues added detail to their claims. They presented evidence of fossils from Tanis—including stunningly well preserved bones, skin, and footprints from what's probably a *Triceratops*—that suggest dinosaurs were indeed witnesses to the asteroid that ushered them out of existence.

"This is a good example that shows a healthy ecosystem in the Dinosauria prior to the extinction," says Sean Gulick, a geophysicist at the University of Texas, Austin, who attended the talk.

Many scientists agree the dinosaurs lived until the asteroid impact, but some paleontologists are reserving judgement on the new data from Tanis. The meeting presentation is "a step in the right direction," says Kay Behrensmeyer, a vertebrate paleontologist at the Smithsonian Institution's National Museum of Natural History. But although the claims "are dramatic and compelling, especially to nonpaleontologists," she says the team hasn't fully considered alternative interpretations that might complicate dating of the fossils.

The fateful asteroid that struck what is today Mexico's Yucatán Peninsula near the town of Chicxulub Pueblo left a crater 150 kilometers in diameter. It sent violent shock waves across the globe, kicked up plumes of hot dust and ash, and triggered volcanic eruptions and roiling tsunamis. The apocalyptic spectacle marked the end of the Cretaceous period, during which dinosaurs had dominated the landscape, and the start of the Paleogene era.

But despite its reputation as a "dino killer," a few paleontologists have raised the possibility that the dinosaurs were already extinct or in serious decline when the asteroid struck. They say dinosaur fossils are rare in deposits 3 meters or so immediately below the thin line of clay, rich in the metal iridium—abundant in asteroids—that marks the impact in rock outcrops around the world. These scientists suggest nonbird dinosaurs had perished 1 million or so years before the impact, perhaps in the wake of bursts of volcanism.

In 2019, DePalma, then a paleontology graduate student at the University of Kansas, Lawrence, and colleagues published a paper in the *Proceedings of the National Academy of Sciences* (*PNAS*) on the Tanis site. They reported finding glassy debris formed by the intense heat of the asteroid's impact buried directly alongside and inside of fossilized fish. Tanis, in the Hell Creek formation, was once located on a riverbank, and the researchers claimed the fossil assemblage represented a burial ground for animals swept up the river valley and killed by surging waters in the minutes and hours following the impact. The paper said nothing about dinosaurs, however.

A few days before the *PNAS* paper went online, *The New Yorker* published an article about the site, describing fossils of dinosaur skin and feathers not mentioned in the *PNAS* paper. In the magazine article, DePalma and colleagues said this suggested dinosaurs were thriving until the asteroid's impact. But some paleontologists, noting that they couldn't evaluate evidence given only in the magazine, professed skepticism. (DePalma says earlier drafts of the *PNAS* paper had included dinosaur data that were removed during later editing rounds.)

Last week, more than 2 years after that kerfuffle, DePalma, now working toward a Ph.D. at the University of Manchester, and colleagues finally presented dinosaur data. One talk, presented by DePalma's colleague Riley Wehr, a geosciences graduate student at Florida Atlantic University, described fossilized bones and a piece of remarkably well preserved bumpy skin identified as likely belonging to a *Triceratops*. These fossils came from the same layers that appeared to record the sedimentary surge that buried the fish. The partially decayed skin indicated the animal had died a few weeks before the impact, Wehr noted.

He also described 11 feathers in the surge deposit that he said were larger than those of any known bird from the Hell Creek formation. "The feathers either came from a dinosaur or a currently unknown bird taxon," he said.

Wehr presented some evidence not mentioned in *The New Yorker*: In a layer of clay immediately below the impact layer, the researchers report several fossilized footprints they say most likely came from an infant *Triceratops*, infant and adult footprints from hadrosaurs, and several less-distinctive tracks attributed to the broad group of carnivorous dinosaurs called theropods. "The dinosaur trackways at Tanis do represent a rich and taxonomically diverse dinosaurian assemblage which immediately precedes the impact," Wehr said in his talk. They reveal "a thriving dinosaurian community," he added.

DePalma says there's no reason to believe these dinosaurs were hanging on long after dying out elsewhere. "It would really shock and surprise me if the Hell Creek formation was [the] only, last isolated nugget of vibrant life before the impact occurred." A paper describing the dinosaur finds is in review at a journal and a few others are being prepared for submission, he adds.

After the haze of skepticism that followed the *New Yorker* article, Gulick says he's glad to see the research presented to a scientific audience, with peer-reviewed papers on the way. He says he reviewed the original Tanis paper in *PNAS* and that "some of what didn't end up in that article … was not quite ready for primetime …[and] needed more work. Now more work has been done and has started to come out and that's great."

The proximity of the fossils and footprints to the Cretaceous-Paleogene boundary "indeed supports the idea that dinosaurs were still present in North America at the time of the Chicxulub impact," says Philippe Lambert, an astrogeologist and director of a research institute that studies an impact crater at Rochechouart in France.

But Behrensmeyer, an expert on fossil preservation, says the researchers haven't shown enough evidence the tracks were made shortly before the impact. The floodwater surge may have eroded away a layer of sediment, revealing older trackways in an underlying rock layer. She thinks DePalma and colleagues have more work to do. "Because of early polarization caused by *The New Yorker* article, it will take a hypothesis-testing approach and extra attention to detail in future manuscripts … to convince the community," she wrote in an email.

Steve Brusatte, a paleontologist at the University of Edinburgh, says the evidence from Tanis is "remarkable," but he, too, is waiting to see whether the dating of the various fossils, feathers, and trackways holds up. "If the dating is solid, then this is a remarkable story, and it proves that dinosaurs were thriving in western North America right up until the asteroid fell from the sky." ■

THE NEW NORMAL

Pandemic spotlight brings attention—and plenty of heat

Scientists like Devi Sridhar leaped to fill an information gap

By **Kai Kupferschmidt**

COVID-19 catapulted Devi Sridhar to a stardom she could scarcely have imagined. Two years ago, she was an accomplished academic, studying disease outbreaks in Brazil, India, and Haiti and running the Global Health Governance Programme at the University of Edinburgh. Sridhar, who is 37, was media savvy; she had co-authored a book with Chelsea Clinton and often spoke on global health. But as COVID-19 exploded, the spotlight shone far brighter: She began advising the Scottish government on pandemic policies, appearing regularly on U.K. morning TV shows, and writing a weekly column for *The Guardian*. She has a book on the pandemic coming out next spring and has amassed more than 300,000 followers on Twitter.

Sridhar is one of dozens of researchers, many in fields like epidemiology, virology, and global health that previously drew little public interest, who became famous almost overnight when they took to social media to comment on pandemic policies and science. "He's real and he's spectacular!" cheered Savannah Guthrie, co-host of *The Today Show*, in reference to a legendary *Seinfeld* episode, after interviewing Ashish Jha, dean of the Brown University School of Public Health. In Germany, virologist Christian Drosten has four times as many Twitter followers—800,000—as the likely next leader of his country, Olaf Scholz. "For the tin foil hats I'm an annoyance, for everyone else the Chuck Norris of science," rocked the skate punk band ZSK, in a tribute song to Drosten.

These scientist-personalities are navigating their fame with a mix of enthusiasm, anxiety, exhaustion—and uncertainty about what their future professional lives will look like. "The anonymity of just being able to kind of do my stuff is gone," says Sridhar, who's been greeted by a Twitter follower in a yoga class.



ILLUSTRATION: KATTY HUERTAS

Sridhar found her voice early in the pandemic and didn't hesitate to use it. "Part of my job is speaking truth to power. And the UK govt is (in my view) getting it wrong," she tweeted in March 2020, after Prime Minister Boris Johnson announced a strategy to "take it on the chin" rather than try to contain the virus. In November 2020, she told a House of Commons committee that the public's desire for a normal Christmas was wishful thinking: "I have to speak bluntly. The virus does not care if it is Christmas."

The information gap Sridhar and others moved to fill was carved partly by months of poor communication from many governments and public health agencies, says Lawrence Gostin, director of the O'Neill Institute for National and Global Health Law at Georgetown University. "They've made so many misjudgments in terms of health communication," he says. Sridhar can "explain what we know, what we don't know, and what we're doing to find out what we don't know," says Gostin, who has known her for years. She "understands the science and she's able to convey it in a simple, clear, honest way." Many of Gostin's British friends listen to Sridhar over their own government, he says.

If the pandemic handed some scientists a megaphone, it also brought a heap of new challenges and worries. "On one hand, I'm very privileged to have such a large platform, and at the same time it's a complete and total exhausting shitshow," says Angela Rasmussen, a virologist at the University of Saskatchewan, whose Twitter following exploded from 250 people to 250,000. For one, scientists can easily get entangled in pandemic politics and suffer the consequences. Sridhar, for example, advises (and frequently lauds) the Scottish government led by the pro-independence Scottish National Party. That has made her a favorite target for those in England who oppose an independent Scotland. She worries whether her directness will jeopardize her ability to secure funding in the future. "You're challenging the establishment," she says. "These are the people who are sitting on all the grant panels in London."

David Fisman, an epidemiologist at the University of Toronto who has been a vocal critic of some pandemic policies in Canada, says colleagues' responses have strained his professional relationships. "I don't think my rank ordering of colleagues, in terms of who I hold in high esteem and who I hold in less esteem, looks anything 2 years into the pandemic like it did at the outset," he says. University of Toronto pharmacologist David Juurlink put it succinctly in a recent tweet: "I miss the days when doctors on Twitter would disagree without referring to each other as human garbage."

The wider public can be even more savage. Like many researchers voicing strong opinions, especially women, Sridhar has received vicious online messages from those opposing masks, lockdowns, and vaccines. "Are you an evil nazi f---ing bitch by any chance, asking for a friend??" a Twitter user wrote one day, after she'd endorsed vaccinating children against COVID-19. "Whenever I open Twitter, I'm so overwhelmed by how angry it is," Sridhar says. She tries to think of her online persona as an avatar onto which people project their fears and anger and frustration. "It's a different universe, there's different rules," she says. "It's not real."

Still, she was unprepared for the hate. "In public health, we've always been the good guys," she says. Growing up in Miami as the daughter of two physicians from India, Sridhar's interest in health intensified around the age of 12, when her father was diagnosed with cancer. (He died a few years later.) At 18, she became the youngest person ever in the United States to receive a Rhodes Scholarship to study at the University of Oxford, where she completed a Ph.D. in anthropology, focusing on the link between malnutrition and infectious diseases in India.

There is something relentlessly sunny about her demeanor, and Sridhar has retained a cheerful irreverence over the years—she once wore flip-flops at Buckingham Palace—and through COVID-19 times. In response to the recent hate message demanding to know whether she was a Nazi, Sridhar retweeted the attack and answered: "Last time I checked … I wasn't. Hope you'll let your friend know! I do like green smoothies, Nigella's chocolate cake, Scotland's gorgeous beaches. In case your friend is wondering." She ended with a hug emoji.

What will happen to these now-household names when the pandemic subsides? Rasmussen plans to keep tweeting about viruses, "but I do hope there's occasion for more tweets about obscure *Beverly Hills, 90210* trivia" too.

Michael Bang Petersen, a political scientist at Aarhus University, predicts many celebrity scientists will continue to speak out. The attention may prove a career boost, he notes, helping scientists secure promotions and funding. Rasmussen, for example, says she found a new job and new collaborations through Twitter. "Without a doubt my own work has been transformed permanently."

Outspoken scientists also have public postures to defend, Petersen says. "In the beginning, everyone was just trying to figure out what is going on," he says. "Now, scientists are more and more locked into a particular position." He worries that will only increase, especially if scientists try to hold on to their followers after the pandemic. "For example, if you have, as a scientist, built up a following on being slightly skeptical about some of the vaccines," he says, "then you may continue that focus just with other types of vaccines."

Gostin hopes trust in public health agencies will return, potentially dimming the spotlight on now-famous scientists. "There will always be room for these independent voices of reason, but I hope in the future they will be amplifying and explaining public health agency messages rather than contradicting them," he says.

For her part, Sridhar hopes to fade into the background. And if another pandemic strikes? "I have a great group of younger researchers. I'll probably pass the baton to one of them," she says. "I feel like I've done my tour of duty." ■

BIOMEDICAL RESEARCH

Should the next NIH director have an M.D.?

Some observers say White House should drop that tradition to broaden the candidate pool

By **Jocelyn Kaiser**

As the search for the next director of the U.S. National Institutes of Health (NIH) gears up, observers are wondering: Is it time for the White House to abandon tradition and choose a biomedical scientist who does not hold a medical degree?

It's been a topic of discussion since NIH Director Francis Collins announced on 5 October he will step down by the end of the year—and urged that a woman succeed him. The question may be especially relevant this time because insisting on a physician at the helm of NIH could complicate efforts to recruit a woman or a person of color (POC). Among potential candidates whose names quickly popped up was University of California, Berkeley, biochemist Jennifer Doudna, who won a Nobel Prize in 2020 for co-discovering the CRISPR gene-editing tool but does not hold an M.D.

There is no legal requirement that the NIH director be a physician, but all 16 directors in the agency's 134-year history have had a medical degree. Collins himself, an M.D. and geneticist with a chemistry Ph.D., has an active medical license. He has occasionally seen patients at the NIH Clinical Center, which treats thousands of people each year, although not as part of patients' official care teams, an NIH spokesperson says.

Jeremy Berg, former director of the National Institute of General Medical Sciences (and former editor-in-chief of *Science*), argued on Twitter that "medical training matters" because the NIH director oversees clinical programs and the Clinical Center. In addition, not having the M.D. title after the NIH director's name "could be a political liability as issues arise," he wrote. Others disagreed: An M.D. requirement "holds back on the NIHs[sic] evolving mission," which includes areas such as public health and bioengineering, tweeted Georgia Institute of Technology biomedical engineer Krishnendu Roy. "The right, politically savvy person can be an asset."



NIH Director Francis Collins (center) with former directors and fellow M.D.s Elias Zerhouni (left) and Harold Varmus.

PHOTO: NATIONAL INSTITUTES OF HEALTH

One former NIH director sees it both ways. Cancer biologist Harold Varmus, an M.D. who stopped practicing medicine in the early 1970s and let his license lapse in the late '80s, a few years before he became NIH director in 1993, says an M.D. is not needed to oversee grantmaking or set up research programs. "The vast majority of the work can be done perfectly well by someone with a Ph.D."

But Varmus says knowing "the vocabulary of medicine" can be "useful." Members of Congress might expect an NIH director to have an M.D., he adds. It could make it "easier" to persuade lawmakers that NIH needs more funding "not just to pursue a basic understanding of science, but to improve health," Varmus says. The degree can also be relevant when members of Congress ask NIH for medical advice for themselves or a family member, Kathy Hudson, a former NIH deputy director, told *STAT*.

Yet others suggest times have changed. "The tradition seems dated," says longtime NIH observer Tony Mazzaschi, chief advocacy officer of the Association of Schools and Programs of Public Health. With NIH's budget now pushing past $41 billion, he thinks management and communication skills "are more essential than ever. Perhaps those skills need to be considered on par with scientific expertise."

Mary Woolley, president of the group Research!America, which advocates for biomedical research funding, agrees that an NIH director's administrative skills should carry a lot of weight. She notes that hospital CEOs were once expected to have medical degree, but that is no longer the case. "I just think it's time to think more broadly," Woolley says.

Woolley and Varmus both served on a committee of biomedical research leaders that recently pondered the M.D. question. The panel, which included past NIH Director Elias Zerhouni and two other former NIH officials, noted in a report last year: "It is essential that the Director has wide knowledge about medicine. However, we do not view the MD degree as an essential attribute."

An M.D. requirement could shrink the pool of high-profile POC and women researchers—ruling out not just Doudna, but also others, including biochemist Erin O'Shea, president of the Howard Hughes Medical Institute, one of the largest private funders of biomedical research in the United States, and neurobiologist Cori Bargmann, head of science at the Chan Zuckerberg Initiative, another large funding philanthropy. (NIH has had just one woman as a director, Bernadine Healy.)

The White House has said it expects to nominate an NIH director before Collins steps down. The post requires approval by the Senate health committee, which could take time because of other pressing business, including the expected nomination of a Food and Drug Administration chief. Confirmation also requires approval by the full Senate. But the NIH candidate could begin to learn the ropes as soon as they are named, Varmus says. "If we had a nominee by the end of the year, that would be a triumph." ■

eppendorf & Science

PRIZE FOR NEURO BIOLOGY



Congratulations
to Amber L. Alhadeff, Ph.D.
Assistant Member
Monell Chemical Senses Center
and University of Pennsylvania

Meet the Winner

Eppendorf & Science Prize for Neurobiology
Congratulations to Amber L. Alhadeff on winning the 2021 Eppendorf & Science Prize for her work on the gut-brain control of hunger circuits. Dr. Alhadeff's research has revealed how hunger-sensitive neurons in the brain receive signals from the gastrointestinal tract, and how they influence food intake and other survival behaviors. Her work helps to answer such questions as why we behave differently when we have not eaten, how we know when to stop eating, how foods we eat influence our brain activity, and why we perceive the world differently when we are hungry or full.

The annual US$25,000 Eppendorf & Science Prize for Neurobiology honors scientists, like Dr. Alhadeff, for their outstanding contributions to neurobiological research. Amber L. Alhadeff is the 20th recipient of this international award. Due to coronavirus (COVID-19), the 2021 winner and finalists will be honored at a virtual Prize Ceremony in November 2021.

You could be next to win this prize.
If you are 35 years of age or younger and currently performing neurobiological research, you could be next to win the 2022 Prize. Deadline for entries is June 15, 2022.

Learn more at: eppendorf.com/prize

eppendorf Science AAAS

AAAS® and Science® are registered trademarks of the American Society for the Advancement of Science, USA. Eppendorf® and the Eppendorf Brand Design are registered trademarks of Eppendorf AG, Germany. All rights reserved, including graphics and images. Copyright © 2021 by Eppendorf AG. Photography: Happy Hour Headshot Philadephia.



THE SIMPLEST OF SLUMBERS

Evidence from evolutionarily ancient creatures is revealing that sleep is not just for the brain

By **Elizabeth Pennisi**

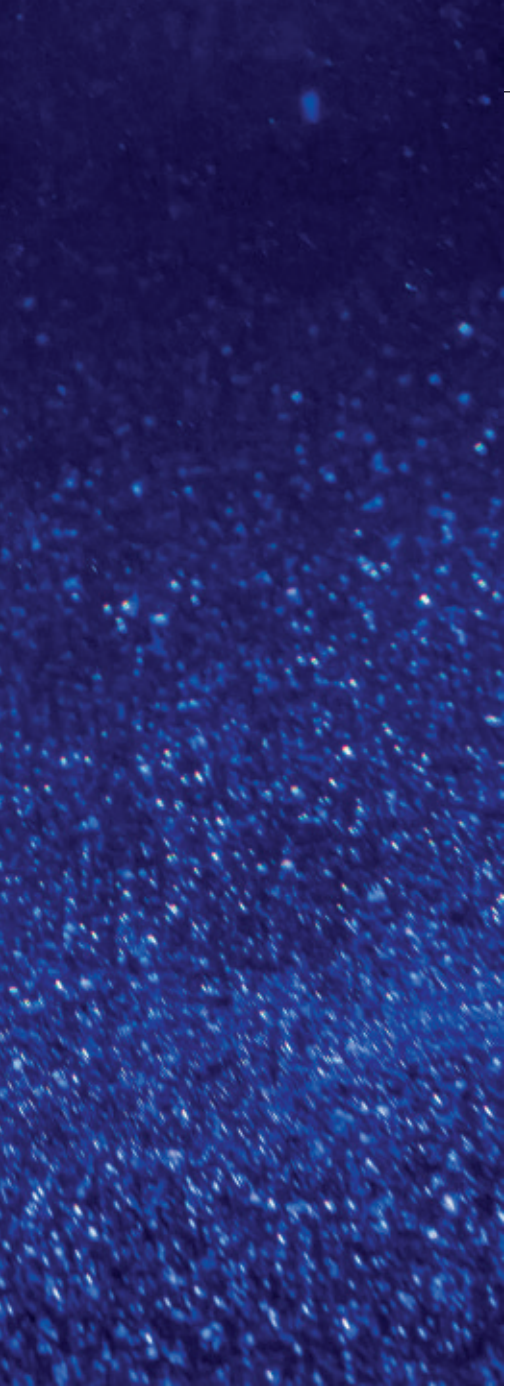
Dive among the kelp forests of the Southern California coast and you may spot orange puffball sponges (*Tethya californiana*)—creatures that look like the miniature pumpkins used for pies. No researchers paid them much mind until 2017, when William Joiner, a neuroscientist at the University of California (UC), San Diego, decided to look into whether sponges take naps.

That's not as silly a question as it seems. Over the past few years, studies in worms, jellyfish, and hydra have challenged the long-standing idea that sleep is unique to creatures with brains. Now, "The real frontier is finding an animal that sleeps that doesn't have neurons at all," says David Raizen, a neurologist at the University of Pennsylvania (UPenn) Perelman School of Medicine. Sponges, some of the earliest animals to appear on Earth, fit that description. To catch one snoozing could upend researchers' definition of sleep and their understanding of its purpose.

Scientists have often defined sleep as temporary loss of consciousness, orchestrated by the brain and for the brain's benefit. That makes studying sleep in brainless creatures controversial. "I do not believe that many of these organisms sleep—at least not the way you and I do," says John Hogenesch, a genome biologist at Cincinnati Children's Hospital Medical Center. Calling the restful, unresponsive state seen in jellyfish and hydra "sleeplike" is more acceptable to him.

But others in the field are pushing for a much more inclusive view: that sleep evolved not with modern vertebrates as previously assumed, but perhaps a half-billion years ago when the first animals appeared. "I think if it's alive, it sleeps," says Paul Shaw, a neuroscientist from Washington University in St. Louis. The earliest life forms were unresponsive until they evolved ways to react to their environment, he suggests, and sleep is a return to the default state. "I think we didn't evolve sleep, we evolved wakefulness."

PHOTO: LANCE HAYASHIDA/CALIFORNIA INSTITUTE OF TECHNOLOGY



One of the first brainless creatures shown to sleep, the upside-down jellyfish, pulses its bell more slowly at night than in the day.

If that's true, sleep in humans, rodents, and other vertebrates is a highly evolved behavior—one adapted to each organism's needs and lifestyle. Gleaning insights into its basic function from those species could be difficult. Earlier evolving creatures, with fewer cell types, less complicated molecular pathways, and simpler behaviors may reveal sleep in its most fundamental form.

So, some sleep researchers have turned to invertebrates such as fruit flies and roundworms—and most recently to sponges and another early-evolving group, placozoans. Already, their work is driving home two key new insights: that sleep's benefits extend far beyond the brain, and that muscles, the immune system, and the gut can all have a say in when and how sleep occurs. That work "might change our focus from studying sleep's role in complex cognitive processes to how it impacts basic cellular function," says Alex Keene, a neurogeneticist at Texas A&M University, College Station.

A new picture of what controls sleep might also lead researchers to new treatments for sleep disorders, says Amita Sehgal, a neuroscientist at UPenn's Chronobiology and Sleep Institute. "The hope is that what we learn will be relevant to understanding why some people can't sleep and also how disrupted sleep might affect their health and performance."

THE EARLIEST STUDIES of sleep defined it by how it changes human behavior: We lie down, close our eyes, remain motionless, and become oblivious to the outside world. Also obvious are the consequences of skipping sleep: We lose our ability to function, struggling to focus in a meeting or dozing off at the wheel.

By the 1950s and '60s, researchers were converging on a definition of sleep based on polysomnography, a combined measure of brain activity, eye movement, and muscle tone that became a gold standard. Neuroscientists figured out how to capture brain activity from electrodes on the surface of the head and discovered that human sleep has two major stages: rapid eye movement (REM), a more active stage in which dreaming occurs; and non-REM, defined by slow, synchronous waves of electrical firing.

Behavioral and physiological tests have revealed how varied sleep can be in the animal world. Cows and other large grazing mammals sleep standing up. Some marine mammals sleep while swimming and some seabirds catnap while flying, letting one half of the brain doze while the other keeps working. Bats sleep about 20 hours a day; wild elephants as few as two. Most of the animals studied with electrical recording techniques have at least two stages of sleep, though the brain activity characterizing these stages can vary. The color changes of the octopus as it sleeps suggest it, too, has several sleep stages.

By the turn of the 21st century, evidence of sleep outside mammals prompted researchers to start to work down the animal tree of life to evolutionarily older species. They had to confront the question of how to define sleep in these simpler species. A sleeping jellyfish looks a lot like a waking one, after all, and is nearly impossible to outfit with electrodes. Researchers must instead recognize where and when simple creatures seek respite and find a behavior that ceases when they sleep. Studies must also poke or otherwise bother the animals to make sure they are unresponsive—and to see whether being forced to stay awake has consequences.

In 2017, Michael Abrams and two other California Institute of Technology graduate students devised such tests for *Cassiopea*, known as the upside-down jellyfish because it tends to stay near the shallow sea floor, pulsing with its tentacles pointing up so more light reaches the photosynthetic microorganisms it relies on for energy. The students observed that at night, this motion slowed from 60 pulses per minute to 39.

To see whether the jellyfish were really asleep, they built a false bottom to the aquarium and lowered it—essentially "pulling the rug out" from under the creatures. At night, the groggy jellyfish were slower to react and swim to the new bottom than in the day. And when the team disturbed the jellyfish by pulsing currents of water over them, the animals were less active the next day—as if having to recuperate from sleep loss. Finally, the drug melatonin, an over-the-counter sleep remedy, slowed their pulses to nighttime speeds. All this without a real brain: Jellyfish have a ring of nerve cell clusters around the rim of their bells.

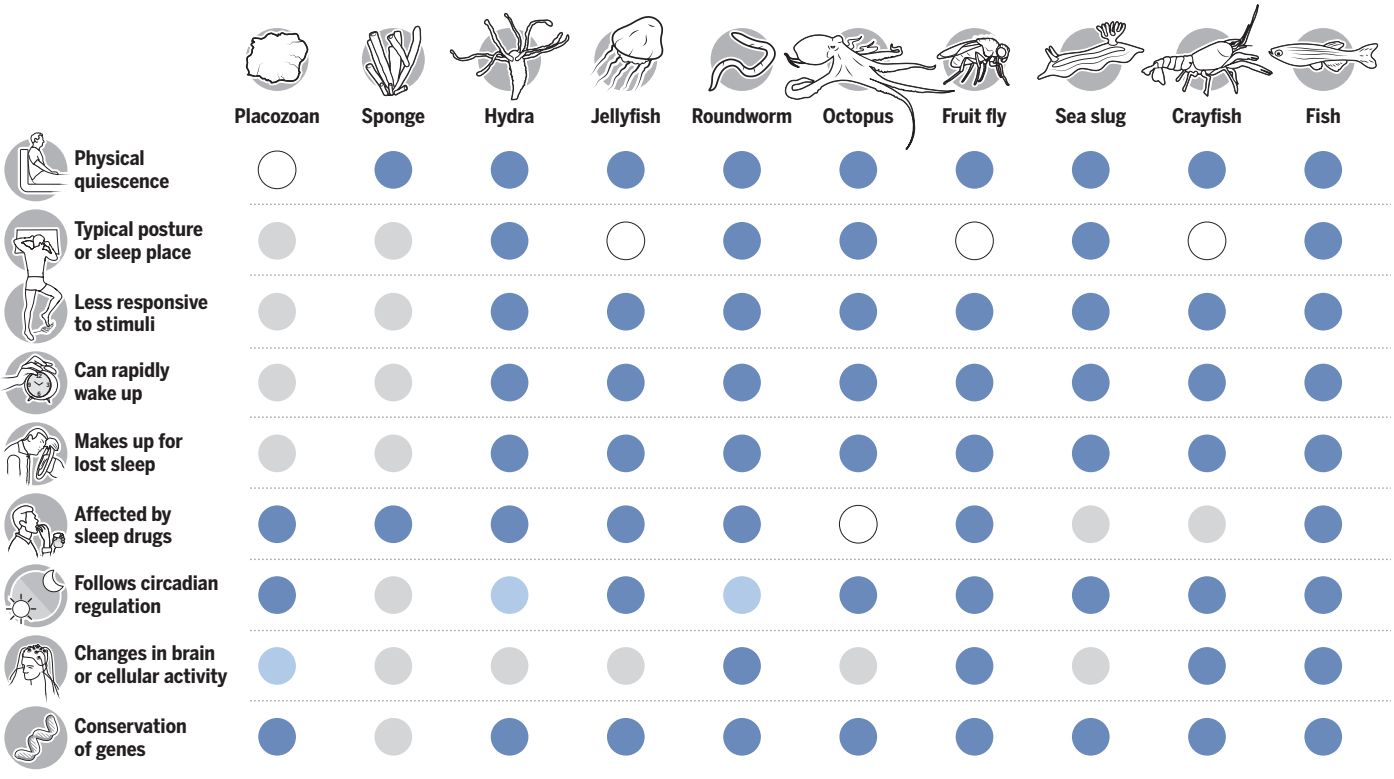
Recently, researchers caught another brainless creature napping: *Hydra vulgaris*, a stationary freshwater relative of jellyfish. Taichi Itoh, a chronobiologist at Kyushu University, and colleagues filmed these centimeter-long animals as they wiggled their tentacles during 12-hour periods of light and dark in the lab. In the dark, the hydra were less active. Other researchers probing sleep in simpler animals have also adopted definitions based on behavioral changes such as reduced responsiveness.

More recently, however, a few are advocating a shift to molecular criteria such as whether an organism has genes that are part of sleep-promoting pathways in mammals and other species known to sleep. For example, Itoh's team reported last year that more than 200 genes changed their activity in sleep-deprived hydra. Several of these genes are involved with sleep in fruit flies, they noted.

"We are moving from a behavioral or physiological definition to a cellular and molecular definition," says Philippe Mourrain, a neurobiologist at Stanford University. "As we define more and more what sleep is [on those levels], we will have an idea of its function."

The signs of sleep

By looking broadly for behaviors characterizing sleep in humans and other organisms, researchers are finding that most animals, even very simple ones, have a restful state. How well each creature satisfies these criteria is controversial, but the work is expanding our understanding of the role and control of sleep, even in humans.



THERE'S NO QUESTION that sleep benefits the brain in creatures that have one (see p. 550). It helps the brain consolidate memories and flush out toxic waste. It may also help the brain stay plastic by pruning and strengthening connections between nerve cells.

But if animals without brains need sleep, those functions can't be the whole story, Sehgal says. "Given that sleep is so widely conserved, it likely serves a fundamental function to preserve basic physiological processes."

Some recent clues from brainless animals suggest sleep factors into energy balance and metabolism. Raizen's team has found the much-studied roundworm *Caenorhabditis elegans* only naps when metabolic demands are high. The larvae go limp for 1 or 2 hours when they are molting and replacing their exoskeletons, or when excess heat or ultraviolet (UV) light causes stress. An enzyme called salt-inducible kinase 3 provides a direct link between sleep and metabolism. Known to help regulate sleep in mammals, it also mobilizes fat stores in *C. elegans* to boost the worm's energy levels. In hydra, too, Itoh's team has found a gene that both regulates metabolism and influences sleep.

Deprivation studies also point to sleep's metabolic role. Sehgal has found fruit flies carrying a mutation that reduces their sleep are less able to metabolize nitrogen, which means they may have trouble breaking down and rebuilding proteins and processing waste. The result is a buildup of charged molecules called polyamines that can damage DNA and RNA, Sehgal's team reported on 2 October on bioRxiv. "When we are sleep-deprived, it's not just brain function that is affected," she says.

Sleep deprivation also appears to attack the gut in fruit flies and mice, by leading to a buildup of harmful molecules known as reactive oxygen species, Harvard Medical School biologist Dragana Rogulja and colleagues reported last year in *Cell*. Somehow, that accumulation leads to early death in both species, the team found. Rogulja suspects the gut, among the first organs to evolve in multicellular animals, was one of the original beneficiaries of sleep, and that "a lot of [sleep's] additional roles evolved as animals become more complex."

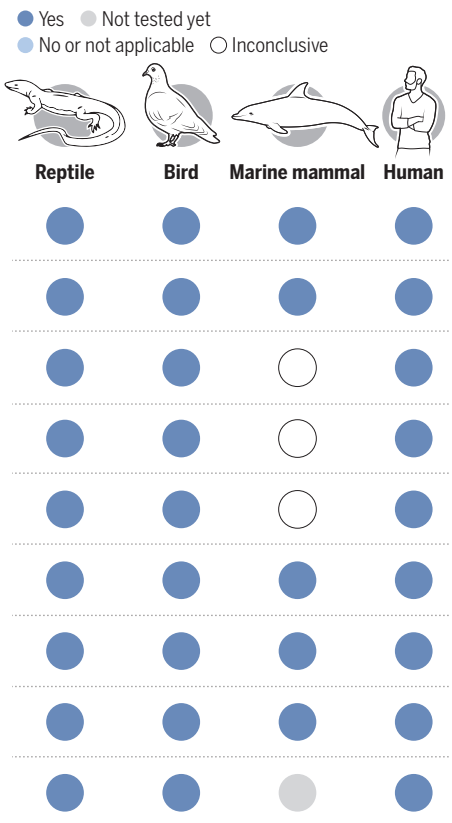
To get to the essence of why animals sleep, however, requires studying it in species so simple they don't even have a gut. Raizen decided to look at placozoans—round, flat, transparent creatures no bigger than a sesame seed that have just two layers of cells, each outfitted with whiplike projections called cilia. Placozoans lack nerve cells; their cells communicate via chemical secretions that control cilia movements. Outside of parasites that live attached to other life forms, placozoans "are the simplest animals on Earth," says Carolyn Smith, a neuroscientist at the National Institute of Neurological Disorders and Stroke who has studied them for more than 10 years.

Placozoans use their cilia to crawl randomly along rocks at the tideline until



The see-through fish *Danionella translucida* may help reveal how the body controls sleep.

CREDITS: (PHOTO) JAMES JAGGARD, PHILIPPE MOURRAIN, ADAM DOUGLASS, AND ADRIADNE PENALVA; (GRAPHIC) N. DESAI/*SCIENCE*



they detect microalgae and stop to graze. They slow down at night, notes Bernd Schierwater, an evolutionary biologist at the University of Veterinary Medicine Hannover. He thinks that slowdown represents "the first [evolutionary] step toward sleep—getting a rhythm for rest" to recharge for the next feeding cycle. That may be enough of a respite for an animal that lacks energy-hungry nerve cells, he says.

Smith thought the idea of a sleeping placozoan was silly—until the jellyfish and hydra studies drove home that sleep was not just for brainy creatures. At times, placozoans rotate in place, which Smith suspects may also represent a kind of sleep. Because these creatures cringe when exposed to UV light, it might be possible to test whether they become unresponsive to UV light in this state, she notes. She provided Raizen with some animals to run such tests. But he couldn't keep them alive in the lab—they are very finicky eaters—and eventually he gave up.

Joiner had similar problems with his puffball sponges. He and his collaborator, marine biologist Greg Rouse of the Scripps Institution of Oceanography, babied the sensitive creatures, which require a constant supply of fresh seawater laden with the microbes they eat. Joiner stopped by the ocean daily on his way to work to the collect seawater. The researchers stuck tubs housing the sponges in an incubator to control light levels and temperature. Then they mounted a digital camera over the tubs to watch for subtle contractions in the sponges' bodies as they pumped water through their chambers to filter out food.

Eventually, with the help of magnetic stirrers in the tubs, they were able to keep the sponges healthy enough that they started to contract—about once every 3 hours. That was exciting, Joiner says, because it gave the researchers a reliable behavior they could monitor for sleep-related changes.

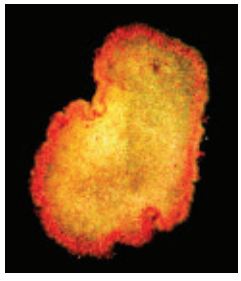
Studies in another sponge species have suggested the animals have cycles of rest that may allow them to reorganize and rejuvenate their cells after pumping water equivalent to 1000 times their volume daily. Marine biologists Sandie Degnan and Bernard Degnan at the University of Queensland, St. Lucia, saw hints of daily rhythms in the contractions in *Amphimedon queenslandica*, found on the Great Barrier Reef. And after sequencing its genome—the first for a sponge—they reported in 2017 that a few genes associated with circadian rhythm in other organisms turned on and off in 24-hour cycles. In unpublished work, graduate student Davide Poli has observed that select regions of the sponges' bodies seem to stop pumping throughout the day, as if working in shifts. That might be "behavior that can be approximated to sleep," he says.

Future studies could use glutamate or other substances that stimulate contractions to keep sponges pumping nonstop for hours or even weeks to test whether their health declines, says Sally Leys, a marine biologist at the University of Alberta, Edmonton. "That could indicate that by being a multicellular animal, you depend on having periods during which the tissues can repair and regenerate."

IF CELLS THROUGHOUT the body benefit from sleep, it makes sense that those cells would have some say in when sleep happens. And the search for sleep's far-flung control switches could lead researchers to new treatments for sleep disorders, which affect 60 million people in the United States alone.

Paul Ketema, a neuroscientist at UC Los Angeles, has studied Bmal1, a ubiquitous protein that regulates gene expression and is known to help sleep-deprived mice stay awake. Until now, researchers had assumed the brain made and used Bmal1 for that task. But Ketema and his colleagues discovered sleep-deprived mice depend instead on Bmal1 produced in the muscles. He suspects the protein may be part of a pathway that somehow helps link muscle exertion to levels of sleepiness. And he's hopeful a Bmal1 drug targeted to muscles may one day counter the negative effects of all-nighters. "I am one of those people who thought sleep was all about the brain," he laughs. "That is not only an erroneous viewpoint," it's also one many of his colleagues still hold, he says.

Other mouse studies have shown the gastrointestinal tract, the pancreas, and fat tissue generate signaling molecules called neurohormones that appear to affect the onset and duration of sleep. Understanding feedback from these organs to the brain "could suggest new pharmacological approaches, with drugs that target organs other than the brain," Sehgal says.


Researchers want to know whether early-evolving placozoans sleep.

Mourrain's team at Stanford is developing a way to watch that feedback process play out cell by cell in the thumbnail-size, transparent fish *Danionella translucida*. By using fluorescent tags and other markers that track the activity of certain molecules in the fish's brain and body, his team will observe how different types of cells control sleep—and benefit from it—over time.

When Mourrain started to study sleep in fish 15 years ago, "there was a lot of pushback that fish don't sleep," he recalls. That changed about 2 years ago, when he and his colleagues developed polysomnography techniques for fish and discovered that, like people, they cycle between quiet and active sleep states. "It was a tipping point for our field," Mourrain says, convincing skeptics that fish could be good stand-ins for mammals in sleep studies.

Will the orange sponges be the next creatures to stun the sleep skeptics? Maybe not right away. Joiner and Rouse couldn't keep their sponges healthy long enough to carry out reliable experiments. After months of fits and starts, they paused the work to rethink the set up. Then, COVID-19 hit and shut the experiment down. Joiner doesn't have the staff to start it back up.

But another sponge may step in instead to show that even the simplest animals truly sleep. In emails, Itoh coyly referred to his lab's work on these simple creatures. Mum's the word for now, it seems. "These are ongoing projects," he wrote. "Please look forward to them." ■

PHOTO: WOLFGANG JAKOB/SCHIERWATER LAB

INSIGHTS

POLICY FORUM



PHOTO: INDRANIL ADITYA/NURPHOTO/GETTY IMAGES

SOCIAL SCIENCE

Informing sleep policy through field experiments

Evidence is particularly needed from poorer communities

By **Gautam Rao[1], Susan Redline[2], Frank Schilbach[3], Heather Schofield[4], Mattie Toma[1]**

Recent progress in sleep science has greatly improved our understanding of the neurobiology of sleep and its importance for physical and mental functioning. Lab experiments have shown that insufficient sleep causes declines in cognitive and physiological function, and community studies have documented widespread sleep deprivation. Adults in the US sleep just 6.1 hours per night when objectively measured (*1*), well below the 7 to 9 hours recommended by experts (*2*). Evidence is emerging that sleep duration and quality are even lower in developing countries and among the poor in rich countries (*3*–*5*). This has led to predictions that increased sleep would have profound benefits for society, including increased productivity, academic performance, health, and safety (*6*). Why do people not sleep more, given these predicted benefits?

A standard economic model would posit that, when deciding how much time to spend in bed, people weigh benefits of sleeping more against costs, while facing a fixed overall "time budget" (*7*). When viewed through this lens, it becomes evident that we know surprisingly little about the calculus of sleep in people's lives. Field experiments—experimental studies in natural environments—can



A wholesale vegetable seller sleeps in Kolkata, India. Field experiments can improve our understanding of sleep in natural environments.

help measure the impacts of sleep on the outcomes people value, such as time use, financial well-being, health, and happiness, over long periods of time.

Using affordable and accurate wearable devices such as actigraphs to measure sleep, field experiments combine the strengths of lab and community studies. They use controlled, randomized interventions—as in pragmatic clinical trials—to measure the causal effect of improved sleep in natural environments on real-world costs and benefits. One such example is a recent study in India (see the first box) that randomized 452 adults to treatments that encouraged increases in nighttime sleep duration and/or daytime naps over 3 weeks (*5*).

The measurement of ecologically valid costs and benefits through field studies is central to policy decisions. Lab experiments typically adopt as endpoints sleep itself or aspects of cognition, such as sustained attention, which can be precisely and reliably measured across studies. But it is difficult to know how such effects translate into real-world outcomes. For instance, among data-entry workers enrolled in the same sleep study in India, performance in the commonly used psychomotor vigilance task (PVT) correlated only modestly with people's productivity, hours worked, and earnings in their data-entry job (*5*).

Field experiments also study sleep in natural environments, which often differ markedly from lab conditions. This divergence could be particularly large when studying sleep among the global poor, who often struggle with noise, heat, light, mosquitoes, shared sleep spaces, and physical and psychological distress. The costs and benefits of sleep may be quite different in such contexts.

Because field experiments can accommodate larger sample sizes than is feasible in the lab, they can study more modest but also more realistic changes in sleep. Studying such modest increases in sleep may lead to quite different conclusions than lab studies, which instead typically experimentally induce severe sleep restriction. Researchers can thus evaluate scalable and policy-relevant interventions as they would play out in practice. For example, field experiments have been used to evaluate pragmatic policies to improve sleep, such as delayed school start times (*8*) and restricted work shifts among physicians (*9*).

Because they can be conducted over long durations and in natural settings, field experiments can also capture how people adjust their lives in response to changes in sleep. Chronically sleep-deprived people may cope by structuring their workdays differently, by adopting countermeasures such as increasing caffeine intake, or even by selecting into work that is less sensitive to cognitive performance, thus mitigating the impacts of sleep deprivation.

Field studies can also capture the "opportunity costs" of sleep: the reduced time available for other activities such as work, exercise, and leisure. If people value these activities highly enough, they might reasonably decide to set aside less time for sleep despite the cost of fatigue the following morning. Yet these costs are often neglected in the literature. For instance, before the results of the study in India (*5*) were released, 119 experts from sleep science and economics made predictions about the effect of increased night-sleep duration on work performance. The experts on average predicted a 7% increase in hours worked and a 12% increase in work output, presumably because they expected reduced tiredness and higher motivation and cognitive performance. Instead, increased night sleep came at the cost of having less time available to work. Eighty-three percent of experts made predictions outside the 95% confidence interval of the results (*5*), highlighting the need to explicitly measure opportunity costs in future research.

Despite these potential strengths, field experiments have until recently been rare in sleep science. One reason is the historical dif-

Sleepless in Chennai

A field experiment in Chennai, India (*5*), measured sleep in a low-income urban population and evaluated interventions to improve sleep. Using actigraphy, low levels of nighttime sleep duration and efficiency were documented compared to levels observed in rich countries. At baseline, participants sleep on average just 5.6 hours each night (see the first figure), with an average sleep efficiency of only 70%. Seventy-one percent of participants sleep less than 6 hours per night on average. The study features two cross-randomized interventions to increase sleep:

1. A bundle of interventions to increase nighttime sleep, including devices to improve people's home sleep environment, information, and encouragement and/or modest financial incentives to increase sleep.
2. An offer of a daily half-hour nap in the early afternoon in a quiet office.

The nighttime treatments increased nighttime sleep duration by an average of 27 minutes without affecting efficiency and had no significant impact on a host of outcomes (see the second figure). By contrast, naps resulted in significant improvements but also reduced time available to work (see the second figure). The contrasting effects of naps and nighttime sleep may be explained by naps having higher sleep quality or because naps were timed to coincide with the mid-afternoon circadian dip. Deidentified data are publicly available at www.sleepdata.org.

[1]Department of Economics, Harvard University, Cambridge, MA, USA. [2]Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Beth Israel Deaconess Medical Center, and Harvard Medical School, Boston, MA, USA. [3]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, PA, USA.
Email: heather.schofield@gmail.com

ficulty of objectively measuring sleep in the field. Another is that social scientists who conduct policy experiments have neglected the importance of sleep, and experimental sleep scientists have largely prioritized the controlled and standardized outcome measurements in the lab. Advances in actigraphy have enabled sleep scientists to document sleep in the field (*3*, *10*) and have inspired recent field experiments by social scientists interested in sleep (*5*, *11*).

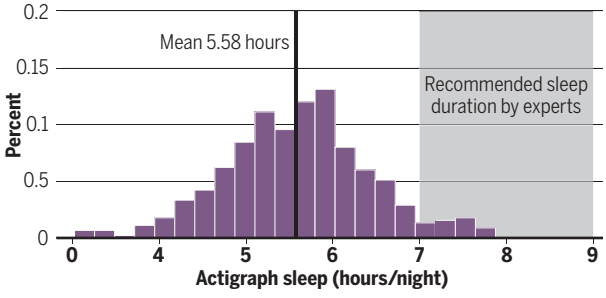
As field research progresses, it should not focus solely on sleep duration but instead should engage with the multidimensional nature of sleep. In addition to time in bed and sleep duration, sleep science has shown that sleep efficiency, fragmentation, and variability are all important dimensions of sleep (*10*). The benefits of increasing sleep duration, say through inducing more time in bed, may be small in settings with low sleep quality unless sleep can be improved along other dimensions as well. Consistent with this view, increased nighttime sleep caused no gains in any outcomes from an intervention that substantially increased sleep duration without improving its (poor) quality (see the first figure). A key objective of future research should be to test interventions that target not only sleep duration, but other dimensions as well. Sleep disorders, such as sleep apnea, are widely underdiagnosed and undertreated, and the impact of untreated sleep disorders in low-income countries requires systematic study.

One such intervention is cognitive behavioral therapy for insomnia (CBT-I). Clinical trials show that a short course of CBT-I improves multiple dimensions of sleep, including sleep efficiency, sleep onset latency, and wake after sleep onset (*12*). In addition to potential benefits from improved sleep quality, CBT-I may have low opportunity costs for recipients because it does not require more time in bed. Importantly for policy, CBT-I can also be effectively delivered as a self-help therapy over the internet (*12*). A brief, manualized version has also recently been developed for low-resource settings where mental-health professionals are scarce (*13*). Existing trials of CBT-I and other clinical sleep interventions have understandably focused on sleep itself as the primary outcome and on a limited set of important secondary outcomes such as positive affect or comorbidities. Adding measures such as time use, earnings, work performance, and financial well-being to such trials would be a valuable next step. In one example of such work, an internet-delivered CBT-I intervention increased job satisfaction and reported self-control among workers (*14*).

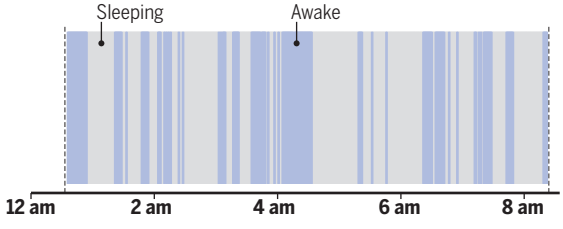
Beyond providing a more complete picture of the costs and benefits of such clinical interventions, what kinds of research would help inform policies on sleep around the globe? Economists have emphasized two broad justifications for policy-makers to take action to change individual behaviors.

Sleep patterns among the urban poor in India

Data are from (*5*), based on 452 study participants in Chennai, India (see the first box). Hours asleep, averaged at the participant-level over 7 nights in their home environments (before any treatments were administered), are shown (top). Sleep-wake patterns are shown for a single night for a representative participant (bottom). This participant stayed in bed for 7 hours and 45 minutes but slept for only 5 hours and 20 minutes (sleep efficiency = 69%). The participant awoke 31 times, and the longest sleep episode lasted 45 minutes. These statistics are close to the median observation on each dimension.



A typical night



First, people may sleep less than is socially optimal because of externalities, where some people are harmed by others' actions. For example, noise pollution due to traffic may disturb sleep in the surrounding community. Similarly, constraints imposed on individuals by others—such as employers or policy-makers—may result in too little sleep if decision-makers undervalue the benefits of sleep. For instance, school-district administrators may underappreciate the value of later school start times, which increase student sleep duration and reduce tardiness and daytime sleepiness (*8*). Or employers may undervalue well-rested employees because some of the benefits will accrue outside the office and may not translate into overall higher work output (*5*). Here, policies that directly set standards, regulate work shifts and school timings, or impose fines on the externality-producing actions may be warranted.

Second, individuals may not be trading off their private costs and benefits of sleep correctly. This might be the case because well-established psychological factors such as limited self-control keep them from following their own best interests, recommending policies that provide incentives to sleep, limit temptations, or help individuals commit to getting enough sleep. Alternatively, individuals may underestimate the value of sleep or simply not have the necessary information to overcome barriers (*11*). In such cases, information campaigns or other forms of encouragement to increase sleep could be warranted. Measurement of people's beliefs about the benefits and costs would be valuable to understand the scope for intervention.

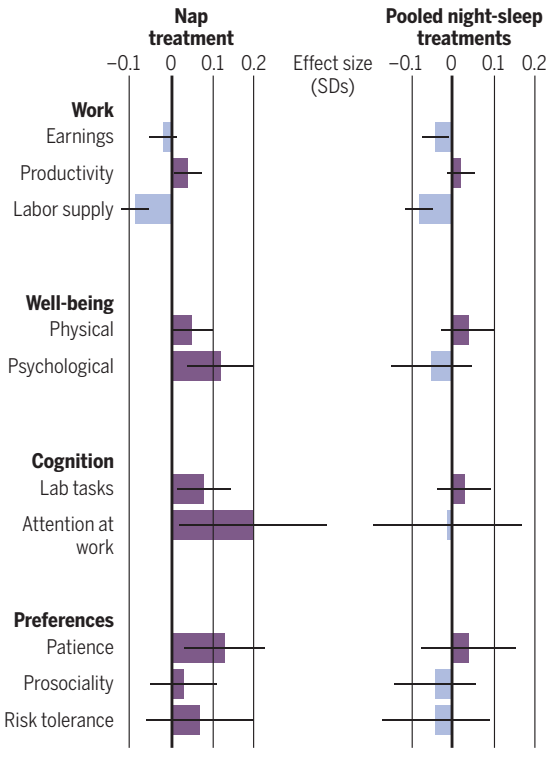
Of course, this view of sleep as a choice has its limitations. Low sleep duration and quality due to psychological distress or uncomfortable living conditions are not freely chosen. But they may still be amenable to actions such as using therapy to manage stress or investing in a better sleep environment. Similarly, sleep may be shortened by hard constraints faced by individuals, such as work shifts and school timings. These may be thought of as choices made by different actors in society—for example, by firms or policy-makers.

In each case, it is crucial to rigorously evaluate proposed policy solutions and iterate on their design. Such research has often exploited quasi-experiments, in which changes in policy are "as if" random. An example is a study that compared students in two schools, one of which delayed school start times on one day of the week by 60 minutes to accommodate a weekly staff meeting (*8*). Policy evaluation—even of quite complex and ambitious policies—is increasingly amenable to randomized field experiments in partnership with policy-makers and organizations. An excellent example is a recent field experiment that evaluated the elimination of extended-duration shifts for resident physicians. Motivated by correlational evidence that sleep-deprived physicians make more mistakes, this study found that eliminating extended shifts of 24 hours or more (and instead limiting shifts to no more than 16 hours) increased physicians' sleep but increased serious medical errors, possibly because of increased patient handoffs between physicians (*9*). This underscores the value of evaluating policy changes using field experi-

GRAPHIC: N. DESAI/*SCIENCE* BASED ON BESSONE *ET AL.* (*5*)

The effects of nighttime and nap treatments

Point estimates and 90% confidence intervals are shown, from the study from the first box (*5*), for the pooled night-sleep interventions (right) and the nap intervention (left). All outcomes are standardized around the control group mean, and a positive value implies a "better" outcome. The following outcomes are index variables, which combine multiple measures: (i) physical well-being: performance in a stationary biking task, blood pressure, and self-reported illnesses, pain, and health; (ii) psychological well-being: self-reported depression, happiness, life possibility, life satisfaction, and stress; (iii) lab tasks: measures of attention, memory, and inhibitory control; (iv) attention at work: attention to piece rates in the data-entry task; (v) patience: savings and present bias; (vi) prosociality: choices in dictator, trust, and ultimatum games; (vii) risk tolerance: negative of choices in risk aversion and loss aversion lotteries.



GRAPHIC: N. DESAI/*SCIENCE* BASED ON BESSONE *ET AL.* (*5*)

High-priority sleep research areas to inform policy-making

- Address the dearth of sleep research—both prevalence studies and field experiments—in developing countries and in low-income settings in high-income countries.
- Include a broader range of outcomes in pragmatic clinical trials, complementing usual measures of physical and mental health with economic, time use, and social outcomes such as earnings, work performance, social relationships, financial decision-making, and other outcomes directly valued by individuals and policy-makers.
- Conduct field experiments to study interventions to improve multiple facets of sleep (duration, satisfaction, regularity, efficiency, and timing) such as CBT-I in general populations.
- Study costs and benefits of naps in everyday lives, including measures of opportunity costs, such as foregone work time, and costs associated with accommodating napping.
- Study behavioral barriers to sleep at the individual level, for instance measuring the importance of incorrect beliefs about the benefits of sleep and the role of behavioral biases such as limited self-control.
- Study concrete social policies in both high- and low-income countries with the potential to improve sleep, such as improved mental health care, changes in school and work timing, noise abatement, housing vouchers, and environmental regulations. Improved sleep could be an overlooked and previously unmeasured benefit of such policies.
- Evaluate the impact of undiagnosed sleep disorders (e.g., sleep apnea) on sleep quality, health, and productivity in low-income settings.

ments, which can identify unintended consequences and guide policy corrections.

The existence of very different barriers to healthy sleep across different contexts and even across individuals implies the need for a diverse set of sleep policy tools. For example, because low-income individuals are at higher risk of housing insecurity and poor mental health (*15*), policies to improve access to housing and mental health care may be particularly valuable for them. Tools to alleviate behavioral factors such as overuse of electronics before bed may instead matter relatively more in high-income settings. Naps could be particularly beneficial in developing countries in tropical settings, where nighttime sleep may be of low quality and the opportunity cost of foregone work in the afternoon (when temperatures peak) may be low. Indeed, offering short naps to workers in India yielded more benefits than extending nighttime sleep (see the second figure). Field research across all these different contexts is needed to guide policy, as costs and benefits of policies are likely to differ substantially across contexts.

We have identified several high-priority sleep research areas to inform policy-making (see the second box). Bodies such as the National Center for Sleep Disorders Research could help to identify initiatives across the US National Institutes of Health and other federal agencies to tackle these questions, and nongovernment funders such as the Wellcome Trust and the Gates Foundation could also provide valuable support. Most pressingly, sleep research has been concentrated in rich countries and in higher-income populations within those countries, even though the majority of the world's population lives in lower-income settings. Informing policy requires documenting how much and how well people sleep across the world and measuring the costs and benefits of improved sleep, including sleep duration but also other dimensions such as efficiency, fragmentation, and variability. More frequent interdisciplinary collaborations between sleep and social scientists to tackle these issues are a natural next step in this research agenda. ■

REFERENCES AND NOTES

1. D. S. Lauderdale, K. L. Knutson, L. L. Yan, K. Liu, P. J. Rathouz, *Epidemiology* **19**, 838 (2008).
2. N. F. Watson *et al.*, *J. Clin. Sleep Med.* **11**, 591 (2015).
3. A. Schokman *et al.*, *Sleep Health* **4**, 543 (2018).
4. W. M. Troxel *et al.*, *Sleep* **41**, zsy140 (2018).
5. P. Bessone, G. Rao, F. Schilbach, H. Schofield, M. Toma, *Q. J. Econ.* **136**, 1887 (2021).
6. M. P. Walker, *Why We Sleep: Unlocking the Power of Sleep and Dreams* (Scribner, 2017).
7. J. E. Biddle, D. S. Hamermesh, *J. Polit. Econ.* **98**, 922 (1990).
8. K. E. Minges, N. S. Redeker, *Sleep Med. Rev.* **28**, 86 (2016).
9. C. P. Landrigan *et al.*, *N. Engl. J. Med.* **382**, 2514 (2020).
10. J. Chung *et al.*, *Am. J. Epidemiol.* kwab232 (2021).
11. M. Avery, O. Giuntella, P. Jiao, "Why don't we sleep enough? A field experiment among college students," IZA Discussion paper 12772 (2019).
12. A. van Straten *et al.*, *Sleep Med. Revi.* 1 (2017).
13. H. E. Gunn, J. Tutek, D. J. Buysse, *Sleep Med. Clin.* **14**, 235 (2019).
14. C. M. Barnes, J. A. Miller, S. Bostock, *J. Appl. Psychol.* **102**, 104 (2017).
15. M. Ridley, G. Rao, F. Schilbach, V. Patel, *Science* **370**, eaay0214 (2020).

ACKNOWLEDGMENTS

We thank I. Frankenthal for research assistance. Time spent on research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health (NIH) under award K01AG055691 and by National Heart Lung and Blood Institute of the NIH under award HLR35135818. S.R. reports grant funding from NIH and Jazz Pharma and consulting fees from Jazz Pharma, Eli Lilly, and Esai Inc. unrelated to this work. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. All authors contributed equally to this work.

10.1126/science.abk2594

PERSPECTIVES

CANCER

Clearing stressed cells

Cell cycle arrest produces a p21-dependent secretome that initiates immunosurveillance of premalignant cells

By **Virinder Reen**[1,2] and **Jesús Gil**[1,2]

Complex organisms repair stress-induced damage to limit the replication of faulty cells that could drive cancer. When repair is not possible, tissue homeostasis is maintained by the activation of stress response programs such as apoptosis, which eliminates the cells, or senescence, which arrests them (*1*). Cellular senescence causes the arrest of damaged cells through the induction of cyclin-dependent kinase inhibitors (CDKIs) such as p16 and p21 (*2*). Senescent cells also produce a bioactive secretome (the senescence-associated secretory phenotype, SASP) that places cells under immunosurveillance, which is key to avoiding the detrimental inflammatory effects caused by lingering senescent cells on surrounding tissues. On page 577 of this issue, Sturmlechner *et al.* (*3*) report that induction of p21 not only contributes to the arrest of senescent cells, but is also an early signal that primes stressed cells for immunosurveillance.

Senescence is a complex program that is tightly regulated at the epigenetic and transcriptional levels. For example, exit from the cell cycle is controlled by the induction of p16 and p21, which inhibit phosphorylation of the retinoblastoma protein (RB), a transcriptional regulator and tumor suppressor. Hypophosphorylated RB represses transcription of E2F target genes, which are necessary for cell cycle progression. Conversely, production of the SASP is regulated by a complex program that involves super-enhancer (SE) remodeling and activation of transcriptional regulators such as nuclear factor κB (NF-κB) or CCAAT enhancer binding protein–β (C/EBPβ) (*4*).

SEs are large enhancers enriched in specific chromatin modifications and genes regulated by SEs often modulate cell fate decisions. Sturmlechner *et al.* searched for senescence-associated SEs (SASEs) that were conserved across species and cell types, and were activated by various senescence inducers. One of 11 conserved SASEs was in proximity to *CDKN1A*, which encodes p21. Suppressing p21 expression in mouse and human cells in vitro as well as mouse hepatocytes in vivo allowed senescent cells to reenter the cell cycle while decreasing the expression of multiple SASP components. Indeed, p21-expressing cells can swiftly secrete proinflammatory factors that partially overlap with the SASP. The authors called this the p21-activated secretory phenotype (PASP).

Sturmlechner *et al.* found that activation of p21 following stress rapidly halted cell cycle progression and triggered an internal biological timer (of ~4 days in hepatocytes), allowing time to repair and resolve damage (see the figure). In parallel, C-X-C motif chemokine 14 (CXCL14), a component of the PASP, attracted macrophages to surround and closely surveil these damaged cells. Stressed cells that recovered and normalized p21 expression suspended PASP production and circumvented immunosurveillance. However, if the p21-induced stress was unmanageable, the repair timer expired, and the immune cells transitioned from surveillance to clearance mode. Adjacent macrophages mounted a cytotoxic T lymphocyte response that destroyed damaged cells. Notably, the overexpression of p21 alone was sufficient to orchestrate immune killing of stressed cells, without the need of a senescence phenotype. Overexpression of other CDKIs, such as p16 and p27, did not trigger immunosurveillance, likely because they do not induce CXCL14 expression.

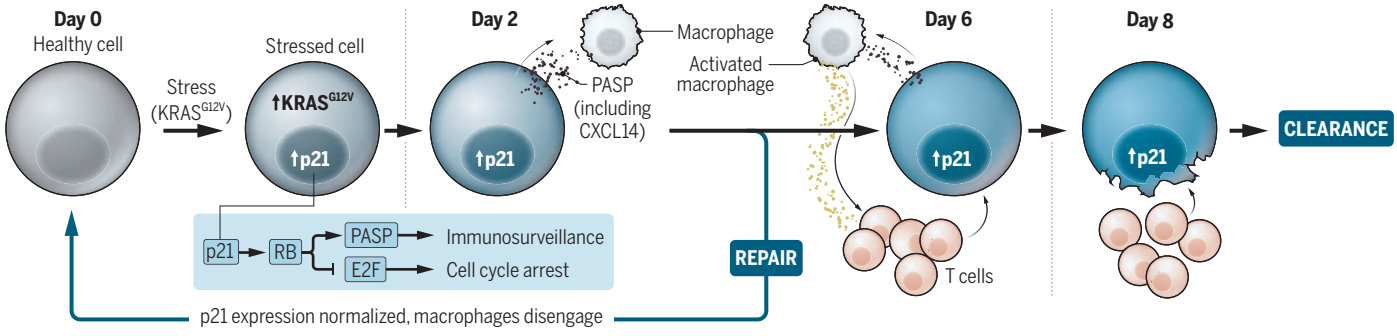
In the context of cancer, senescent cell clearance was first observed following reactivation of the tumor suppressor p53 in liver cancer cells. Restoring p53 signaling induced senescence and triggered the elimination of senescent cells by the innate immune system, prompting tumor regression (*5*). Subsequent work has revealed that the SASP alerts the immune system to target preneoplastic senescent cells. Hepatocytes expressing the oncogenic mutant $NRAS^{G12V}$ (Gly^{12}→Val) become senescent and secrete chemokines and cytokines that trigger $CD4^+$ T cell–mediated clearance (*6*). Despite the relevance for tumor suppression, relatively little is known about how immunosurveillance of oncogene-induced senescent cells is initiated and controlled.

The mutant $KRAS^{G12V}$ oncoprotein induces mitogenic stress that increases p21 expression, initiates senescence, and triggers immunosurveillance. Sturmlechner *et al.* show that $KRAS^{G12V}$ expression in mouse hepatocytes can attract macrophages and subsequently cytotoxic T cell clearance in a p21-dependent manner. Upon p21 ablation in hepatocytes,

[1]MRC London Institute of Medical Sciences, London, UK. [2]Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK. Email: jesus.gil@imperial.ac.uk

Policing stressed cells

Cell stress (such as that elicited by oncogenic $KRAS^{G12V}$) induces p21 activation, which can orchestrate cell cycle arrest and immunosurveillance through retinoblastoma (RB) hypophosphorylation. p21 induction triggers the p21-associated secretory phenotype (PASP), including secretion of C-X-C motif chemokine 14 (CXCL14), which attracts macrophages. p21 initiates a biological timer that gives stressed cells time to repair damage and normalize p21. If repair is not possible, the timer expires: Macrophages are activated and recruit T cells to clear stressed cells.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

these stressed cells fail to promote immunosurveillance, and small clusters of proliferating $KRAS^{G12V}$-expressing premalignant cells become apparent in vivo.

In addition to controlling the cell cycle, p21 is a regulator of vital cellular processes including apoptosis, DNA repair, and cell motility (*7*). Sturmlechner *et al.* now find that p21 is responsible for placing stressed cells under immune control, which prevents the earliest stage of neoplastic transformation. Strategies that can enhance immune infiltration within the tumor microenvironment are of considerable interest. Therefore, PASP or CXCL14 induction could be explored as a potential therapeutic approach to augment anticancer immune responses. In this context, the repercussions of prolonged PASP-mediated immune stimulation on T cell exhaustion would be worth investigating. Conversely, studies have reported chronic p21 expression in the liver to be a marker of poor prognosis for liver cancer (*8*), raising questions about how the role of p21 and PASP changes once tumors have been established.

The induction of senescence in mutant KRAS-driven tumors in the lung or the pancreas elicit different immune responses (*9, 10*). Whether the findings of Sturmlechner *et al.* extend beyond the liver will need to be examined. In addition, the SASP induces senescence in neighboring cells in a paracrine manner (*11*). Whether in the right context the PASP can also cause paracrine growth arrest of adjacent cells is currently unclear. Another conundrum is the expression (or lack thereof) of p53, which often occurs in tumors. Previous studies have reported that continuous p21 expression in p53-null cells can paradoxically be oncogenic, resulting in DNA replication deregulation and genomic instability (*12*). Whether p21-dependent immune clearance can be reactivated in p53-null tumors requires investigation. Such future studies should allow for therapeutic exploitation of stress-related immunosurveillance for a variety of senescence-associated diseases, such as cancer, fibrosis, and beyond. ■

REFERENCES AND NOTES

1. V. Gorgoulis *et al.*, *Cell* **179**, 813 (2019).
2. N. Herranz, J. Gil, *J. Clin. Invest.* **128**, 1238 (2018).
3. I. Sturmlechner *et al.*, *Science* **374**, eabb3420 (2021).
4. N. Tasdemir *et al.*, *Cancer Discov.* **6**, 612 (2016).
5. W. Xue *et al.*, *Nature* **445**, 656 (2007).
6. T. W. Kang *et al.*, *Nature* **479**, 547 (2011).
7. N. A. Warfel, W. S. El-Deiry, *Curr. Opin. Oncol.* **25**, 52 (2013).
8. S. Marhenke *et al.*, *Gut* **63**, 1501 (2014).
9. M. Ruscetti *et al.*, *Cell* **181**, 424 (2020).
10. M. Ruscetti *et al.*, *Science* **362**, 1416 (2018).
11. J. Birch, J. Gil, *Genes Dev.* **34**, 1565 (2020).
12. P. Galanos *et al.*, *Nat. Cell Biol.* **18**, 777 (2016).

ACKNOWLEDGMENTS

J.G. has been a consultant for Unity Biotechnology, Geras Bio, Myricx Pharma, and Merck KGaA. J.G. owns equity in Geras Bio and is a named inventor in MRC and Imperial College patents related to senolytic therapies.

10.1126/science.abm3229

CORONAVIRUS

Defective viral RNA sensing linked to severe COVID-19

Genetic variation in a sensor of double-stranded RNA can exacerbate COVID-19

By **John Schoggins**

Why do some people with COVID-19 get sicker than others? Maybe exposure to a particularly high dose of the causative virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), accounts for the difference. Perhaps deficiencies in diet, exercise, or sleep contribute to worse illness. Although many factors govern how sick people become, a key driver of the severity of COVID-19 appears to be genetic, which is common for other human viruses and infectious agents (*1*). On page 579 of this issue, Wickenhagen *et al.* (*2*) show that susceptibility to severe COVID-19 is associated with a single-nucleotide polymorphism (SNP) in the human gene 2′-5′-oligoadenylate synthetase 1 (*OAS1*).

The authors reasoned that SARS-CoV-2 should be inhibited by interferon-mediated antiviral responses, which are among the first cellular defense mechanisms produced in response to a viral infection. Interferons are a group of cytokines that induce the transcription of a large cadre of genes, many of which encode proteins with the potential to directly inhibit the invading virus. Wickenhagen *et al.* interrogated many hundreds of these putative antiviral proteins for their ability to suppress SARS-CoV-2 in cultured cells and found that OAS1 was particularly potent against SARS-CoV-2.

OAS1 is an enzyme that is activated in the presence of double-stranded RNA, which is scattered along an otherwise single-stranded SARS-CoV-2 genome because of an assortment of RNA hairpins and other secondary structures. Once activated, OAS1 catalyzes the polymerization of adenosine triphosphate (ATP) into a second messenger, 2′-5′-oligoadenylate. This then triggers the conversion of ribonuclease L (RNaseL) into its active form so that it can cleave viral RNA, effectively blunting viral replication (*3*). Wickenhagen *et al.* found that OAS1 is expressed in respiratory tissues of healthy donors and COVID-19 patients and that it interacts with a region of the SARS-CoV-2 genome that contains double-stranded RNA secondary structures (see the figure).

OAS1 exists predominantly as two isoforms in humans—a longer isoform (p46) and a shorter version (p42). Genetic variation dictates which isoform will be expressed. In humans, p46 is expressed in people who have a SNP that causes alternative splicing of the *OAS1* messenger RNA (mRNA). This results in the utilization of a terminal exon that is not used to translate p42. Thus, the carboxyl terminus of the p46 OAS1 protein contains a distinct four–amino acid motif that forms a prenylation site. Prenylation is a posttranslational modification that targets proteins to membranes. In cell culture experiments, Wickenhagen *et al.* showed that only OAS1 p46, but not p42, could inhibit SARS-CoV-2. However, when the prenylation site of p46 was engineered into p42, this chimeric p42 protein was able to inhibit SARS-CoV-2, which strongly implicates a role for OAS1 specifically at membranes.

Why are membranes important? SARS-CoV-2, like all coronaviruses, co-opts cellular membranes at the endoplasmic reticulum to form double-membrane vesicles, in which the virus replicates its genome. Thus, membrane-bound OAS1 p46 may be specifically activated by RNA viruses that form membrane-bound vesicles for replication. Indeed, the unrelated cardiovirus A, which also forms vesicular membranous structures, was inhibited by OAS1. Conversely, other respiratory RNA viruses, such as human parainfluenza virus type 3 and human respiratory syncytial virus, which do not use membrane-tethered vesicles for replication, were not inhibited by p46.

Wickenhagen *et al.* examined a cohort of 499 COVID-19 patients hospitalized in the UK. Whereas all patients expressed OAS1, 42.5% of them did not express the antiviral p46 isoform. These patients were statistically more likely to have severe COVID-19 (be admitted to the intensive care unit). This suggests that OAS1 is an important antiviral factor in the control of SARS-CoV-2 infection and that its inability to activate RNaseL results in prolonged infections and severe disease, although other factors likely contribute.

Department of Microbiology, UT Southwestern Medical Center, Dallas, TX, USA. Email: john.schoggins@utsouthwestern.edu

The authors also examined animals known to harbor different coronaviruses. They found evidence for prenylated OAS1 proteins in mice, cows, and camels. Notably, horseshoe bats, which are considered a possible reservoir for SARS-related coronaviruses (*4*), lack a prenylation motif in their OAS1 because of genomic changes that eliminated the critical four–amino acid motif. A horseshoe bat (*Rhinolophus ferrumequinum*) OAS1 was unable to inhibit SARS-CoV-2 infection in cell culture. Conversely, the black flying fox (*Pteropus alecto*)—a pteropid bat that is a reservoir for the Nipah and Hendra viruses, which can also infect humans—possesses a prenylated OAS1 that can inhibit SARS-CoV-2. These findings indicate that horseshoe bats may be genetically and evolutionarily primed to be optimal reservoir hosts for certain coronaviruses, like SARS-CoV-2.

Other studies have now shown that the p46 OAS1 variant, which resides in a genomic locus inherited from Neanderthals (*5*–*7*), correlates with protection from COVID-19 severity in various populations (*8*, *9*). These findings mirror previous studies indicating that outcomes with West Nile virus (*10*) and hepatitis C virus (*11*) infection, both of which also use membrane vesicles for replication, are also associated with genetic variation at the human *OAS1* locus. Another elegant functional study complements the findings of Wickenhagen *et al.* by also demonstrating that prenylated OAS1 inhibits multiple viruses, including SARS-CoV-2, and is associated with protection from severe COVID-19 in patients (*12*).

There is a growing body of evidence that provides critical understanding of how human genetic variation shapes the outcome of infectious diseases like COVID-19. In addition to OAS1, genetic variation in another viral RNA sensor, Toll-like receptor 7 (TLR7), is associated with severe COVID-19 (*13*–*15*). The effects appear to be exclusive to males, because *TLR7* is on the X chromosome, so inherited deleterious mutations in *TLR7* therefore result in immune cells that fail to produce normal amounts of interferon, which correlates with more severe COVID-19. Our knowledge of the host cellular factors that control SARS-CoV-2 is rapidly increasing. These findings will undoubtedly open new avenues into SARS-CoV-2 antiviral immunity and may also be beneficial for the development of strategies to treat or prevent severe COVID-19. ■

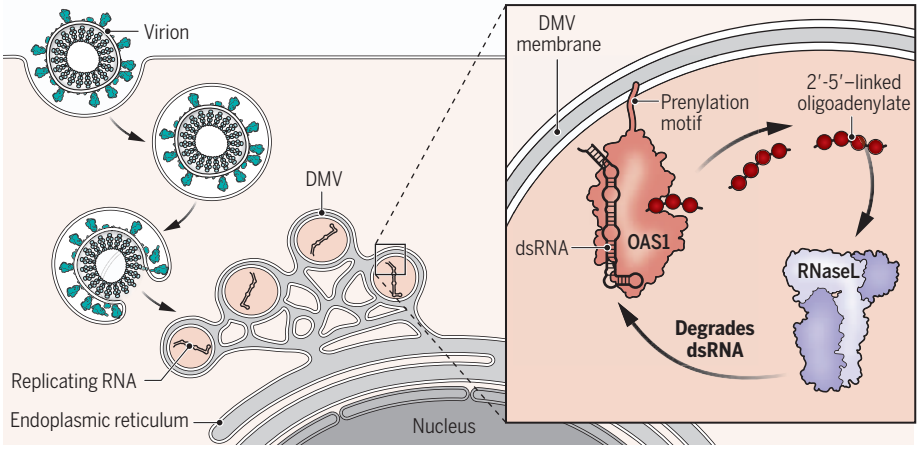
REFERENCES AND NOTES

1. J. L. Casanova, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E7118 (2015).
2. A. Wickenhagen *et al.*, *Science* **374**, eabj3624 (2021).
3. H. Kristiansen, H. H. Gad, S. Eskildsen-Larsen, P. Despres, R. Hartmann, *J. Interferon Cytokine Res.* **31**, 41 (2011).
4. S. Lytras, W. Xia, J. Hughes, X. Jiang, D. L. Robertson, *Science* **373**, 968 (2021).
5. S. Zhou *et al.*, *Nat. Med.* **27**, 659 (2021).
6. H. Zeberg, S. Pääbo, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026309118 (2021).
7. F. L. Mendez, J. C. Watkins, M. F. Hammer, *Mol. Biol. Evol.* **30**, 798 (2013).
8. A. R. Banday *et al.*, medRxiv 10.1101/2021.07.09.21260221 (2021).
9. E. Pairo-Castineira *et al.*, *Nature* **591**, 92 (2021).
10. J. K. Lim *et al.*, *PLOS Pathog.* **5**, e1000321 (2009).
11. M. K. El Awady *et al.*, *J. Gastroenterol. Hepatol.* **26**, 843 (2011).
12. F. W. Soveg *et al.*, *eLife* **10**, e71047 (2021).
13. T. Asano *et al.*, *Sci. Immunol.* **6**, eabl4348 (2021).
14. C. Fallerini *et al.*, *eLife* **10**, e67569 (2021).
15. C. I. van der Made *et al.*, *JAMA* **324**, 663 (2020).

10.1126/science.abm3921

Degrading viral RNA

Once inside a host cell, coronaviruses form double-membrane vesicles (DMVs) to replicate their RNA genome. When anchored to membranes, 2′-5′-oligoadenylate synthetase 1 (OAS1) detects double-stranded RNA (dsRNA) secondary structure and activates ribonuclease L (RNaseL), which degrades viral RNA. The OAS1 p42 isoform lacks the prenylation motif, so it is not membrane bound and RNaseL is not activated, leading to unrestrained viral replication and correlating with severe COVID-19.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

SOCIAL SCIENCE

Documenting Indigenous dispossession

A major data synthesis maps tribal displacements

By **Donald L. Fixico**

Land dispossession through forced migration of Indigenous peoples in the United States has long been studied. But there is fragmented understanding of the traumatic loss of homelands and the personal depth of how Native people have struggled to adjust to confined home spaces, federally reserved areas known as reservations. On page 578 of this issue, Farrell *et al.* (*1*) describe and analyze a first-of-its-kind large dataset that can tell us much about how each displaced tribe was removed from one area to another. The authors drew on sources of information that have been known for many years but that had not been integrated because of the enormous task involved. This macro analysis enables us to quantitatively reexamine, in comprehensive ways, what Indian Country looked like as it was displaced and diminished and how the tribal land base might be viewed today (see the figure).

There are nearly 600 federally recognized tribes in the United States today and 326 reservations. The idea of an Indian reservation dates to 1658, when the General Assembly of the British Virginia Colony passed an act and the Mattaponi of the Powhatan confederation of tribes agreed to live on a reservation. The first official removal involved the Delaware, when their leaders signed a treaty in 1818. Reservations (some in California are known as rancherías) in the United States were created through treaties, congressional laws, and presidential executive orders. The first treaty reservation was created by the United States for the Oneida and Tuscarora tribes with the Treaty of Fort Stanwix in 1784.

To help comprehensively chronicle resettlements, Farrell *et al.* incorporated data from sources such as public archives;

Department of History, School of Historical, Philosophical, and Religious Studies, Arizona State University, Tempe, AZ, USA. Email: donald.fixico@asu.edu

tribal publications; government correspondence with tribes; judicial records during the Indian Claims Commission years (1946–1978); Interior Department records since 1824, including maps; and 374 ratified United States–Indian treaties.

The Farrell *et al.* dataset locates tribes in finite spaces during the 19th century, allowing better understanding of tribal movements and investigation of the reasons for their involuntary or voluntary movements before the government's forced migrations under the Indian Removal Act signed into law by President Andrew Jackson in 1830 (*2*). The massive dataset covers a long range of time, from the early 1820s, and includes a second major episode of tribal land loss due to the Dawes Act of 1887. This legislation undermined traditional communal land tenure practices by allotting individual plots of what had been tribal land as private property to each enrolled member of a tribe. An estimated 173 tribes, bands, and communities were involved in land allotments (*3*). This left large surpluses of land and contributed to drastic reductions in the land area owned by Native Americans. From a pre-Columbus estimate of 5 million Native people in North America, the population dropped below 238,000 by the end of the 1800s, when Indians were called the "vanishing race" (*4*).

With this new dataset, it is possible to learn through stages or phases how this happened rather than guesswork with cartography. Imagine a hypothetical map of the "United Indian Nations of America", and this research informs us of the location of tribes after their forced migrations, and in some instances of multiple migrations such as the Delaware Nation of the New England area, who were removed nine times and now reside within the Cherokee Nation in Oklahoma. One of the problems that the research had to solve was dealing with the fluctuating boundaries of Indigenous communities. In addition, tribal domains sometimes consisted of two or more groups in alliances occupying the same hunting or fishing area, such as the Sauk and Fox, Cheyenne and Arapaho.

The "termination policy" that ended trust status protections in 109 cases of Indigenous lands, and the relocation program from 1952 and 1953 to 1975, led to the demand from energy companies seeking coal, oil, gas, and uranium. Without the government's trust protection, companies dispossessed Native individuals and tribes of their lands and resources. Increased demand for tribal natural resources in the 1960s and 1970s forced those tribes with natural resources to form a protective organization called the Council of Energy Resource Tribes (CERT) in 1975 with 25 original member tribes (*5*). Natural resources of timber, oil, gas, coal, and uranium exist in substantial quantities on only about 60 reservations today. The majority of tribal lands today have a very small portion of land with oil-producing and gas-producing wells.

By comparing postremoval reservations in the 19th century with current reservations, with consideration to climate change exposure and natural resources on tribal lands, the research tells us that tribal economic development, especially in the area of agriculture, is hindered by extreme heat and decreasing annual precipitation. Eighty of 220 tribes saw an increase in drought severity, and 105 saw an increase in wildfire risk. There are 310 forested reservations in 24 states, for example, but about 70% of all US wildfires estimated in 2020 included burning on tribal lands (*6*).

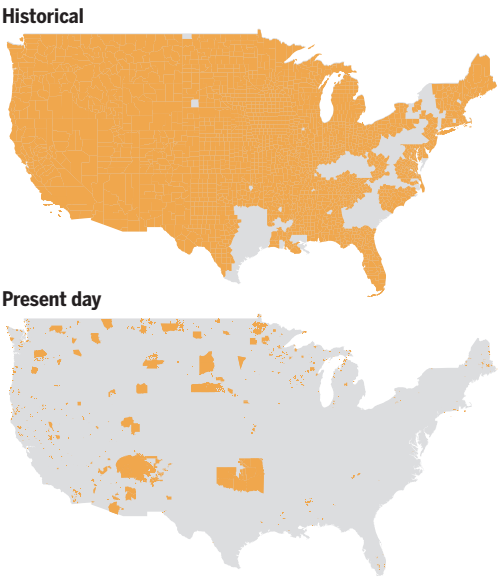
Objectivity is one of the challenges facing this research because most of the data sources have been produced by non-Indigenous parties. Without sufficient Native perspectives, there is a continual general bias in studying United States–tribal relations. As the first and most accurate macroscopic quantitative dataset at the present, the work by Farrell *et al.* is an important development. Possible next steps could include refining the dataset to include more Native input data, such as descriptions of resources on reservations from the Indian-Pioneer History Project for Oklahoma (1861–1936) (*7*) and the Doris Duke Indian Oral History Collections developed in the 1960s and 1970s (*8*).

With human population increases, decreasing natural resources, and climate change, these data need to be shared widely, and research partnerships formed, with organizations such as the National Congress of American Indians (*9*). The Inter Tribal Council of Arizona of 21 tribes, Inter-Tribal Council of the Five Civilized Tribes in Oklahoma, and other Native organizations could use these data and partner with researchers in planning toward sustaining natural resources. Such cooperation may include working at the international level with the Indigenous Environmental Network, founded in 1990, to bring awareness of environmental injustice on tribal lands.

The basin area of Nevada and Utah or the Four Corners area of the Southwest, where natural resources have been heavily extracted, could benefit from regional focus and updated estimates to help inform developing tribal economies under climate change conditions. About 243 of nearly 600 tribes are involved in the gaming industry, but it is a myth that all of them are making huge profits (*10*). Such research and partnerships could help them to develop and diversify their economies and sustain their remaining resources so that they would be less dependent on gaming. The need for such a dataset is justified and would be most beneficial to various agencies of the federal government, Native organizations, and especially the tribal nations (*11*). ■

A sharp decline in tribal land coverage

Farrell *et al.* estimate the historical (top) and present-day (bottom) aggregate land base for all tribes (●).



GRAPHIC: N. DESAI/*SCIENCE*, BASED ON FARRELL *ET AL.* (*1*)

REFERENCES AND NOTES

1. J. Farrell *et al.*, *Science* **374**, eabe4943 (2021).
2. J. Ostler, in *Surviving Genocide: Native Nations and the United States from the American Revolution to Bleeding Kansas* (Yale Univ. Press, 2019), pp. 247–375.
3. F. E. Hoxie, in *A Final Promise: The Campaign to Assimilate the Indians, 1880–1920* (Univ. Nebraska Press, 1984), pp. 1–39.
4. B. W. Dippie, in *The Vanishing American: White Attitudes & U.S. Indian Policy* (Univ. Press Kansas, 1982), pp. 3–45.
5. D. L. Fixico, in *The Invasion of Indian Country in the Twentieth Century: American Capitalism and Tribal Natural Resources* (Univ. Press Colorado, 1998), pp. 159–163.
6. Congressional Research Service, Wildfire statistics, 6 August 2021; https://sgp.fas.org/crs/misc/IF10244.pdf.
7. Indian Pioneer Papers Collection; https://digital.libraries.ou.edu/whc/pioneer/.
8. The Doris Duke Indian Oral History Collection at the University of Oklahoma, Norman; https://digital.libraries.ou.edu/whc/duke/.
9. The National Congress of American Indians; www.ncai.org.
10. D. L. Fixico, in *Indian Resilience and Rebuilding: Indigenous Nations in the Modern American West* (Univ. Arizona Press, 2013), pp. 170–178.
11. K. Frantz, in *Indian Reservations in the United States: Territory, Sovereignty, and Socioeconomic Change* (Univ. Chicago Press, 1993), pp. 156–273.

10.1126/science.abl6288

NEUROSCIENCE

Memories off the top of your head

Cortical layer 1 has a special role in long-term memory

By **Jiyun N. Shin[1,2], Guy Doron[3,4,5], Matthew E. Larkum[4,5]**

Memories are generally considered to be embodied in the connection strengths between neurons. Finding the exact synaptic connections responsible for memories out of hundreds of trillions at first appears impossible. But new research may be taking us a step closer. Specific brain areas under the neocortex are known to be required for memory formation (including the hippocampus and basal ganglia), while long-term memory relating to facts and knowledge (semantic memory) is thought to reside in the neocortex (*1*). However, the precise location of semantic memory is still unknown, preventing researchers from understanding how and where memories are formed.

Accumulating evidence in rodents suggests that neocortical layer 1, the outermost layer of the brain, might be a key site for long-term plasticity (*2*) and learning (*3*, *4*). Common to all these studies is the observation that subcortical memory structures predominantly target this layer. For instance, higher-order thalamic input targets layer 1 in the sensory cortex, where it has also been associated with learning (*5*) and enhanced activity in the tuft dendrites of pyramidal neurons, the principal excitatory neurons of the neocortex (*2*). Similarly, it was shown that projections from the amygdala to layer 1 are crucial for fear memory formation (*6*). More recently, it was shown that the hippocampus also targets neocortical layer 1 via parahippocampal structures and is crucial for associative learning (*7*). This leads to the hypothesis that layer 1 is the locus of semantic memory formation and storage in the neocortex, where the synapses encoding memories are likely to be found.

Layer 1 of the neocortex is enigmatic, and there is still no consensus about its function. It stands out as the only layer almost devoid of cell bodies, which makes it even more surprising that it is the target of so many long-range neuronal axon fibers. These fibers tend to come from higher cortical areas and higher-order thalamic nuclei and are known to convey context-related feedback information (*8*) and correlate with learning (*5*, *9*, *10*). Layer 1 consists of a dense complex of overlaying dendrites made up of the tops of pyramidal neurons (tuft dendrites). Thus, it is this part of the pyramidal neurons that receives feedback information, whereas the bottom of these neurons primarily receives feature-related, feed-forward information. Therefore, the pyramidal neuron is ideally positioned to link the two information streams (*8*).

The question remains: Why would memory-related structures target context in layer 1 rather than features encoded in the lower layers? Adding memory to an existing cognitive theory describing the role of cortical dendrites in perceptual experience may provide the answer (*11*). This theory of perception suggests a functional definition of "feature" and associated "context," where a feature is defined as the primary output of a cortical module ("column"), and context is defined as relevant information from across the brain that arrives in layer 1 of that column. Thus, for an "orange column" responding during a perceptual experience, "shape" would be a part of the context for this column. This perceptual theory explains how the properties of pyramidal neurons bind context to features.

Linking this theory to memory could explain why and how memory structures influence layer 1 of the neocortex. There are two possible scenarios: the memory content scenario, in which memory structures projecting to layer 1 directly provide additional context; or the memory switch scenario, in which memory structures modulate or gate other context-related inputs to layer 1. In both scenarios, experiences would involve the short-term association of context with features, and memory would involve encoding and stabilization of this association. For example, recognizing a tiger should evoke distributed neuronal activity around the cortex according to the features accessible via the senses and the context derived from previously learned associations (see the figure). Seeing a tiger for the first time would involve temporary associations that would need to be formed or stabilized.

This implies that associative memory is the convergence of different types of information from various cortical and subcortical sources to layer 1 and is therefore intrinsically distributed, even in a given cortical area. It is possible that the different memory-related brain structures projecting to layer 1 provide different criteria for stabilization of context association (novelty detection, importance, emotional valence, or deviation from expectation, etc.). Overall, semantic memory is hypothesized to be the long-term association of different contexts with particular features in neocortical layer 1.

The two scenarios entail certain predictions. The memory content scenario predicts that layer 1 inputs constitute memory content themselves, show long-term plasticity, and require the source structure to be activated during retrieval. Evidence for the memory content scenario has been observed directly for thalamic projections to layer 1 (*5*). Conversely, the memory switch scenario explains how memory retrieval could occur without the source structure. A recent study showed that learning to associate a stimulus with a reward requires hippocampal input to layer 1 via the medial temporal lobe structures but could subsequently be retrieved without this input (*7*).

It is possible that the two scenarios could operate in parallel. For example, it is well established that the hippocampus mediates spatiotemporal context that is an important aspect of episodic autobiographical memory (*1*). Thus, inputs to layer 1 could provide both a criterion for consolidating associated context (memory switch scenario) and a constitutive part of the context itself (memory content scenario), providing the autobiographical component by linking time and location to memories.

The precise mechanisms of memory stabilization in layer 1 remain unknown. Stabilization may involve the up- or down-regulation of synaptic inputs to the densely packed tuft dendrites of pyramidal neurons in layer 1 (*9*), or possibly more-complex circuit refinements involving heterosynaptic plasticity induction including other cortical layers (*2*). There is also evidence that local inhibition mechanisms in layer 1 can shape and modulate long-range contextual inputs (*5*). Layer 1 inhibitory neurons are optimally positioned to determine synaptic plasticity via calcium-dependent signaling in the pyramidal cell tuft dendrites. Indeed, several classes of inhibitory neurons reside in or influence layer 1 (*8*). These inhibitory sources

[1]New York University Comprehensive Epilepsy Center, New York, NY, USA. [2]Department of Neurology, New York University School of Medicine, New York, NY, USA. [3]Scientific and Competitive Intelligence, R&D, Pharmaceuticals, Bayer AG, Berlin, Germany. [4]Institute for Biology, Humboldt University of Berlin, Berlin, Germany. [5]NeuroCure Cluster of Excellence, Charité – Universitätsmedizin Berlin, Berlin, Germany. Email: guy.doron@bayer.com; matthew.larkum@hu-berlin.de

may have complementary roles and competing influences on the associative properties of pyramidal neurons, with some controlling dendritic calcium spikes through direct dendritic inhibition and others controlling the firing of inhibitory neurons to release dendrites from inhibition (disinhibition). Moreover, inhibition in layer 1 has itself been found to be plastic and undergo experience-dependent changes (*3*, *8*). Such modulation may form the basis of selection and activation of engram pyramidal cortical neurons—the neurons whose changes in firing encode new memories (*12*).

Conversely, stabilization could also involve the regulation of postsynaptic dendritic excitability (*4*). Modulation of the intrinsic excitability of the tuft dendrites could, in principle, have a large effect on the output of the neuron because they generate powerful local dendritic *N*-methyl-D-aspartate (NMDA) and calcium spikes. Other open questions include the extent to which local dendrite-specific protein synthesis in the apical dendritic compartment is required for memory stabilization and how the influence of memory-related projections to layer 1 is mediated, that is, via specific neurotransmitters or neuromodulators that may be spatially targeted.

The hypothesis that associative, context-related memories are stored in layer 1 raises interesting questions about the psychology of learning. Features are thought to be learned during early development in a critical window within which the dominant form of learning involves the extraction of regularities (*13*). For example, children typically absorb new motor skills and low-level features, such as the accent of a second language, with great ease. According to our hypothesis, learning features would be independent of layer 1 and the tuft dendrites of pyramidal neurons. Consistently, the complex dendritic properties necessary for binding context and features are unavailable before adolescence (*14*). Feature learning would therefore modify synapses distributed throughout the other layers of the cortical column.

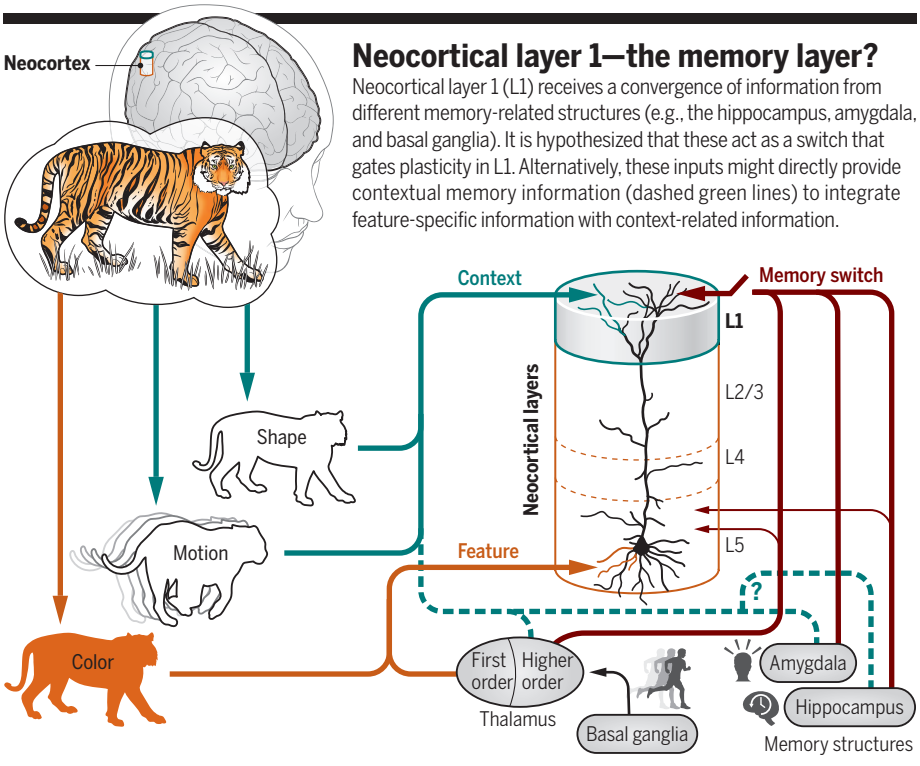
Thus, our proposal is that there are two phases of learning. The first phase involves the grounding of the cortical network in the statistical features most prevalent in the external world, which takes a long time (many repetitions) to establish but is then relatively stable over time. The second phase involves a more dynamic learning of context and high-level concepts that establish an internal model using the putative predictive role of the pyramidal neuron tuft dendrites (*11*). For example, adults can more quickly and explicitly learn the higher-level structure of a second language (e.g., grammar), while never completely acquiring the detailed features of the accent.

The shape of pyramidal neurons could facilitate this two-phase model of learning by allowing plasticity mechanisms to be segregated and separately modulated in the different dendritic trees at the top and the bottom of the pyramidal neuron. The physical segregation of learning within the neuron would also allow independent access to memory of features and associative memory. Furthermore, top-down information could be deployed flexibly without compromising feed-forward information that should remain stably grounded in the outside world. The link between top-down control and semantic memory might explain how it can be explicitly accessed. Accordingly, whether grammatical knowledge is accessed unconsciously versus explicitly would depend on whether the knowledge is encoded as features or context.

Information segregation using multicompartment neurons may also inform machine learning models (*15*). Despite their successes, modern artificial networks still perform very poorly on tasks that require inference from very little information, which is a distinct feature of adult learning in intelligent animals. Identifying the locus of memory in the neocortex allows more biologically inspired models to be generated. For example, it was found that after learning, pyramidal neurons emitted bursts of action potentials that were more salient to memory retrieval (*7*). It is therefore possible that the mammalian brain uses a ternary rather than binary system of outputs (i.e., "0," no output; "1," low-frequency output; or "2," burst output). This could signal the type of information (statistical versus associative) to downstream neurons and facilitate credit assignment for learning in a feedback system (*15*). Such observations can offer inspiration for the design of more robust and intuitive machine learning principles. Therefore, the identification of layer 1 as the locus of semantic memory promises to accelerate our understanding of learning and memory in the human brain and provide insights for the development of treatments for memory disorders as well as the design of architectures for artificial intelligence. ■

Neocortical layer 1—the memory layer?

Neocortical layer 1 (L1) receives a convergence of information from different memory-related structures (e.g., the hippocampus, amygdala, and basal ganglia). It is hypothesized that these act as a switch that gates plasticity in L1. Alternatively, these inputs might directly provide contextual memory information (dashed green lines) to integrate feature-specific information with context-related information.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE* BASED ON THOMAS SPLETTSTOESSER

REFERENCES AND NOTES

1. Y. Dudai, A. Karni, J. Born, *Neuron* **88**, 20 (2015).
2. L. E. Williams, A. Holtmaat, *Neuron* **101**, 91 (2019).
3. E. Abs *et al.*, *Neuron* **100**, 684 (2018).
4. J. Cichon, W.-B. Gan, *Nature* **520**, 180 (2015).
5. M. B. Pardi *et al.*, *Science* **370**, 844 (2020).
6. Y. Yang *et al.*, *Nat. Neurosci.* **19**, 1348 (2016).
7. G. Doron *et al.*, *Science* **370**, eaaz3136 (2020).
8. B. Schuman *et al.*, *Annu. Rev. Neurosci.* **44**, 221 (2021).
9. H. Makino, T. Komiyama, *Nat. Neurosci.* **18**, 1116 (2015).
10. S. Manita *et al.*, *Neuron* **86**, 1304 (2015).
11. M. Larkum, *Trends Neurosci.* **36**, 141 (2013).
12. T. Kitamura *et al.*, *Science* **356**, 73 (2017).
13. C. Blakemore, G. F. Cooper, *Nature* **228**, 477 (1970).
14. J. J. Zhu, *J. Physiol.* **526**, 571 (2000).
15. A. Payeur *et al.*, *Nat. Neurosci.* **24**, 1010 (2021).

ACKNOWLEDGMENTS

Supported by the Deutsche Forschungsgemeinschaft (EXC 257 NeuroCure and SFB 1315 – 327654276) and EU Horizon 2020 (Human Brain Project and European Research Council). We thank B. Rudy, A. Holtmaat, J. Aru, S. Elz, and R. Naud as well as the reviewers for insightful comments and SciStyle for artwork advice. G.D. is employed by Bayer AG.

10.1126/science.abk1859

POLYMER CHEMISTRY

Polymer design for the circular economy

Addition of keto groups to polyethylene helps it degrade while maintaining its properties

By **Margaret J. Sobkowicz**

The field of polymer science and engineering has produced elegant chemical structures and highly functional economical materials during the past century. However, society still faces fundamental challenges recycling even the simplest polymer, polyethylene (PE). The advent of Ziegler-Natta and metallocene catalysts (*1*) has enabled synthesis of PEs with targeted branching structures, spurring their use in an ever-increasing range of applications, including broad adoption in single-use packaging. By comparison, approaches to recover the ethylene monomer or otherwise recycle PE have not kept pace. On page 604 of this issue, Baur *et al.* (*2*) describe a catalytic approach to redesigning PE to facilitate its inclusion in the circular economy by maintaining its properties but making it easier to degrade to monomers.

In 2016, 242 metric tons of plastic waste were generated worldwide (*3*), and this number continues to rise annually. In theory, polymers could be easily remelted and reformed into new objects, but the recycling rate has stagnated at only 9% of all plastics produced globally (*4*), or only 14% of plastic packaging (*5*). This low recycling rate not only represents an inefficiency in manufacturing and loss of material value but also results in unintended leakage to the environment (*6*) and damage to ecosystems (*7*).

At the heart of the struggle is that separations, purifications, and subsequent thermomechanical processing cannot completely restore recycled plastics to their original quality, and the recovered material ends up in lower-value products. Advanced recycling approaches focus on recovering purified chemicals for repolymerization or use in other chemical applications, such as fuels and solvents (see the figure). Processes that are more energy intensive—such as pyrolysis, gasification, and biochemical digestion—can create higher-value product. Hybrid (*8*) and solvent-based (*9*) processes are also attracting interest. Catalysts can target specific bonds for cleavage and lower the energy threshold for polymer deconstruction.

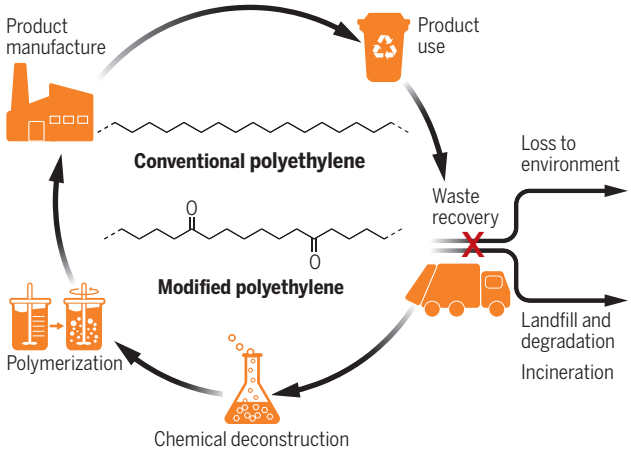
The zero-waste hierarchy (*10*) suggests that the most preferred option for solving the plastics waste problem is to rethink or redesign the polymer itself. Compostable plastics such as poly(lactic acid) could fill this role with appropriate infrastructure (*11*), but their properties are much different than those of the fossil fuel–derived materials to which industry is accustomed. "Oxo-degradable" materials, which contain 1 to 2 wt % of transition metal complexes that promote free-radical chain scission (*12*), may reduce accumulation of visible recalcitrant waste but do not advance recovery and reuse and could inhibit greater adoption of current mechanical recycling and composting infrastructures (*13*).

Baur *et al.* describe a class of nickel-phosphinophenolato catalysts that favor random incorporation of oxygenated carbonyl groups into high–molecular weight PE, rendering it more susceptible to oxidative and enzymatic deconstruction. Notably, they controlled the degree of substitution of keto groups in a nonalternating fashion along the carbon backbone, distinct from other attempts to incorporate such oxygen functionality in the past. These ketone-modified PEs have thermal and mechanical properties comparable to those of conventional PE of similar molecular weight, which is critical if they are to be used as drop-in replacements for existing materials. This approach may give these materials a competitive advantage, considering the existing challenges that have hindered adoption of new biodegradable plastics. The study demonstrates that exposure to ultraviolet light results in photodegradability similar to that of the additive-modified PEs previously described.

This promising approach to polymer redesign should be incorporated into the toolbox of strategies to achieve a circular economy for plastics. Moving forward, the field must develop approaches for processing these materials into packaging structures and work toward controlled depolymerization to recover valuable feedstocks. As with the oxo-degradable additive approach, the degradation products should not be dispersed into the environment unless they are proven safe to ecosystems. More work is needed to establish the time- and environment-dependent breakdown of modified PE to optimize both its in-use properties and end-use recycling methodologies. ■

Circular polymer design

Polyethylene recycling by converting to monomers and chemicals is made easier by Baur *et al.*, who incorporated ketone groups into the backbone chain.



GRAPHIC: N. CARY/*SCIENCE*

Department of Plastics Engineering, University of Massachusetts Lowell, 1 University Avenue, Lowell, MA, 01854, USA. Email: margaret_sobkowiczkline@uml.edu

REFERENCES AND NOTES

1. H. Sinn, W. Kaminsky, in *Advances in Organometallic Chemistry*, F. G. A. Stone, W. Robert, Eds. (Academic Press, 1980), vol. 18, pp. 99–149.
2. M. Baur, F. Lin, T. O. Morgan, L. Odenwald, S. Mecking, *Science* **374**, 604 (2021).
3. K. L. Law *et al.*, *Sci. Adv.* **6**, eabd0288 (2020).
4. J. M. Garcia, M. L. Robertson, *Science* **358**, 870 (2017).
5. United Nations Environment Programme (UNEP), "Single-use plastics: A roadmap for sustainability" (UNEP, 2018).
6. R. Geyer, J. R. Jambeck, K. L. Law, *Sci. Adv.* **3**, e1700782 (2017).
7. A. A. de Souza Machado *et al.*, *Environ. Sci. Technol.* **52**, 9656 (2018).
8. S. Martey *et al.*, *ChemSusChem* **14**, 4280 (2021).
9. T. W. Walker *et al.*, *Sci. Adv.* **6**, eaba7599 (2020).
10. S. Billiet, S. R. Trenor, *ACS Macro Lett.* **9**, 1376 (2020).
11. M. T. Zumstein, R. Narayan, H. E. Kohler, K. McNeill, M. Sander, *Environ. Sci. Technol.* **53**, 9967 (2019).
12. M. Koutny, J. Lemaire, A.-M. Delort, *Chemosphere* **64**, 1243 (2006).
13. New Plastics Economy, "Oxo-degradable plastic packaging is not a solution to plastic pollution, and does not fit in a circular economy" (Ellen MacArthur Foundation, 2019).

ACKNOWLEDGMENTS

I thank D. Kazmer for helpful consultation on this text.

10.1126/science.abm2306





Modern dogs emerged from a population of gray wolves, but when and how remain a mystery.

ANIMAL STUDIES

The question of canine origins

With few solid leads, a paleoanthropologist explores the emergence of the first dogs

By **Greger Larson**

There are roughly 1 billion dogs on Earth, about 10% of which live in homes in North America. For many of us, dogs are not just pets, they are family members with their own bespoke clothing and therapists and trendy snacks (for example, the many varieties of canine pumpkin spice latte treats). Dogs have become so integrated into our lives that it is difficult to imagine how, for at least 90% of our species' history, we lived without them.

In her new book, *Our Oldest Companions: The Story of the First Dogs*, Pat Shipman sets herself two admirable goals. She seeks the complete dog origin narrative, attempting to identify how and when and from where dogs first emerged, and perhaps more ambitiously, to understand "why we have so often misunderstood the full story." To her credit, Shipman is aware of the pitfalls associated with storytelling, including the human bias that enforces an inevitability to the things that are important to us now. Her goal, she explains, is to present the story "the right way around" by questioning fundamental aspects of canine origins that are often assumed and unquestioned.

So what do we know for sure about our dogs' earliest days? We can be certain, for one, that dogs emerged from gray wolves. Although limited amounts of gene flow from other canid species into dogs have been recorded, it was definitely a population of gray wolves from which dogs emerged.

As to when they emerged, that depends on whom you ask. Some scientists have controversially postulated that dogs were with us 32,000 years ago. A minimum bound of 15,000 years is generally accepted. The temporal window of uncertainty results from the difficulties associated with distinguishing the first unambiguous dogs from wolves. Dogs are the result of a fundamental behavioral shift in the relationship between people and wolves, and behaviors are less than easy to infer from either the archaeological or genetic records. In her telling, Shipman presents the case for the older dates but curiously avoids discussing why this deeper end of the time frame has been challenged.

Gray wolves have been distributed across Eurasia and North America for tens of thousands of years, so pinpointing the precise location where dogs first emerged has also proved vexing. This assumes, too, that there was just a single occurrence. It is possible that dogs emerged more than once, independently, from widely geographically separated populations of gray wolves. Shipman avoids choosing a side in this lively debate and leaves the matter unresolved.

The weight of ambiguity begins to take a toll on the book's prose, and Shipman falters by heading down cul-de-sacs and Denisovan tangents. However, she quickly recovers by addressing the twinned questions of how and why dogs came into being. Here, she explicitly states that the domestication of dogs "cannot have been intentional," but she then frequently undermines this position by using the term "domesticate" as a transitive verb, implying that the process was something that people did deliberately to wolves. Shipman avoids stating precisely how she believes dogs first emerged, which is perhaps a braver stance given the lack of evidence for any specific model.

Eager to battle against the dearth of certainty, Shipman heads for Australia, where she suspects a deep investigation of the dingo will offer insights into the dog's hazy origins. Dingoes are a fascinating case study of feralization. Functionally, the way they fill an ecological niche is analogous to wild, Indigenous Australian species, but their genomic ancestry belies a clear-cut relationship to both modern and ancient East Asian domestic dogs and feral New Guinea singing dogs. Undeterred by the overwhelming evidence in support of these relationships, Shipman claims that the ancestors of dingoes were never domestic dogs. Her bewildering position on this issue is likely a symptom of the overall lack of both theoretical and empirical foundations from which we attempt to understand dogs.



Our Oldest Companions: The Story of the First Dogs
Pat Shipman
Harvard University Press, 2021. 272 pp.

In the absence of smoking gun evidence, investigating dog origins can make us feel as though we have no choice but to present myth and contorted logic as fact. Perhaps, instead, there is opportunity. Let us not be intimidated by the abyss. With cleared heads and renewed vigor, we can step forward and collectively seek the origins of our beloved animal companions. ■

The reviewer is at the Palaeogenomics and Bio-Archaeology Research Network, School of Archaeology, University of Oxford, Oxford OX1 3TG, UK. Email: greger.larson@arch.ox.ac.uk

10.1126/science.abl7976

PHOTOS (LEFT TO RIGHT): JASPER DOEST/MINDEN PICTURES; LELUSY/SHUTTERSTOCK

ANTHROPOCENE

Refining the tree of life

Human interventions, both intentional and unintentional, define a new chapter in Earth's storied history

By **Shane Campbell-Staton**

Since humankind's first migration out of Africa, we have trimmed, trained, and reshuffled the tree of life at breakneck speed. In the past 10,000 years, we learned to manipulate the basic building blocks of life through breeding. In the past century, we increased the speed and magnitude of our impact through mass global transportation, industrialized agriculture, and urbanization. In the past decade, we have seen major advances in biotechnology that now allow us to directly rewrite the genetic code of almost any organism to create never-before-seen biological variation.



Popular as pets, fluorescent GloFish were created to serve as sensors.

Our unmatched ability to destroy, reshape, and rebuild life to suit our purposes has allowed our species to become an evolutionary force rivaled only by the five major mass extinctions of Earth's past. Evolutionary biologist Beth Shapiro explores this legacy in her new book, *Life as We Made It*.

In eight chapters, Shapiro takes readers on a succinct and compelling journey through historical events, inventions, and decisions that have forever changed the course of life on Earth. The series of stories she tells effectively combine to outline the rise of humans to global dominance and detail the fates of many species caught in our wake, from the Pleistocene extinctions of North American megafauna to the genetic engineering of glow-in-the-dark fish. Along the way, she shares personal stories of scientific exploration, adventure, and misadventure that have defined her own journey to understand the lasting biological impacts of human activity and the potential of biotechnology to shape the future.

Shapiro details her visits to laboratories, museums, and field sites around the world, introducing readers to the remarkable diversity of extinct and near-extinct species she has studied throughout her career. In chapter one, for example, readers learn about her research as a graduate student in the Canadian Yukon, where she waded ankle-deep through Pleistocene bio-sludge in search of ancient bison remains. Later, in chapter five, we follow her through the ornithology collections of the Oxford University Museum of Natural History as she pursues remnant DNA hidden in specimens of extinct passenger pigeons and dodos.

Many times throughout history, our attempts to correct a course of biological events that we ourselves set in motion have had surprising and unexpected results. In chapter five, Shapiro recounts, for example, how the intentional introduction of several Costa Rican panthers mistakenly thought to be Florida panthers put the endangered Everglades species at risk of being unprotected by the Endangered Species Act, owing to a hesitancy to protect hybrids. Fortunately, an expansion of protections that now preserve distinct lineages, including those that result from human-induced hybridization events, was ultimately approved.

Shapiro's latest literary work makes plain that human intervention will likely be necessary to confront the problems we currently face and that biotechnology has an important role to play across a range of issues, from saving species teetering on the brink of extinction to curing disease to feeding a growing human population. In chapter six, she discusses how genetic engineering can be used to endow endangered species with the tools needed to face novel environmental threats and to produce genetically modified crops that yield more food with fewer resources.

In what is perhaps the greatest accomplishment of *Life as We Made It*, Shapiro clearly articulates key questions whose answers will define how we think about and use the power we now yield. In a world that has been remixed and remade by humans several times over, what is "natural"? When and how should we intervene in evolution? Who should make this call? Throughout the book, she offers personal insights and thoughts that inform these questions, while acting primarily as a facilitator of information that furthers the reader's understanding of current discussions on these topics.

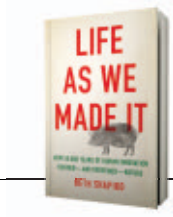
Millions of years from now, our species may be long extinct. However, the microbes, plants, and animals that make up the world after our demise will descend from those species that survive the Anthropocene. What's more, at least some of the genetic code that influences the form, function, and survival of those species may originate from the imaginations of people who exist today.

In a world where we can engineer life in much the same way we engineer skyscrapers, we are bounded more by imagination than by ability. But while buildings will eventually deteriorate in the sands of time, life will keep on ticking. In *Life as We Made It*, Shapiro offers readers a history lesson from which to pull both caution and inspiration. In doing so, she sets the table for a needed conversation about our lasting imprint on the tree of life. It is up to us to take a seat. ■

10.1126/science.abm0633

The reviewer is at the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. Email: scampbellstaton@princeton.edu

Life as We Made It: How 50,000 Years of Human Innovation Refined—and Redefined—Nature
Beth Shapiro
Basic Books, 2021. 352 pp.



SCIENCE & RACE



PODCAST

Algorithms of Oppression: How Search Engines Reinforce Racism
Safiya Umoja Noble
NYU Press, 2018. 256 pp.

It has become increasingly clear in recent years that the internet—long framed as a value-free tool for advancing human knowledge—frequently replicates the flawed societal power structures in which it was created, perpetuating stereotypes and racism under a guise of impartiality. This week on the *Science* podcast, Safiya Noble discusses how discrimination is created and reinforced by search engines. https://scim.ag/3DwULim

10.1126/science.abm5861

PHOTO: PALLY/ALAMY STOCK PHOTO



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

We offer:

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

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

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

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

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

ThermoFisher SCIENTIFIC

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



AN ESTATE GIFT TO THE

American Association for the Advancement of Science

1848 SOCIETY

AAAS

AAAS is dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy, and for the community to speak with one voice to advance science and engineering in the United States and around the world.

You can help us.

By making AAAS a beneficiary of your will, trust, retirement plan, or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

If you plan to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

For more information or to let us know you are including AAAS in your future giving plans, contact the AAAS Office of Philanthropy and Strategic Partnerships at **1-800-215-1969** or **philanthropy@aaas.org.**

Additional details online at **www.aaas.org/1848Society.**



LETTERS

Global fishing subsidies can contribute to overfishing.

Edited by **Jennifer Sills**

WTO must ban harmful fisheries subsidies

Sustainably managed wild fisheries support food and nutritional security, livelihoods, and cultures (*1*). Harmful fisheries subsidies—government payments that incentivize overcapacity and lead to overfishing—undermine these benefits yet are increasing globally (*2*). World Trade Organization (WTO) members have a unique opportunity at their ministerial meeting in November to reach an agreement that eliminates harmful subsidies (*3*). We—a group of scientists spanning 46 countries and 6 continents—urge the WTO to make this commitment.

To curb overfishing, biodiversity degradation and loss, and CO_2 emissions, and to safeguard food and livelihoods, WTO members must prohibit fisheries subsidies that cause harm, such as those that lower the cost of fuel and vessel construction and those that provide price support to keep market prices artificially high (*2*). Subsidies to distant-water fishing fleets must be eliminated to prevent overfishing on the high seas and in waters under national jurisdiction. Such subsidies threaten low-income countries that rely on fish for food sovereignty (*4*, *5*). Exceptions to the rules—known as special and differential treatment—should be considered for small-scale fishers that use low-impact gears or that fish for subsistence, but only if decoupled from incentivizing overfishing (*6*).

An effective agreement must eliminate subsidies for fuel (*7*), distant-water and destructive fishing fleets (*4*, *5*), and illegal and unregulated vessels in line with the aims of Sustainable Development Goal 14.6 (*8*). To ensure accountability, it should also support low-income countries' efforts to meet their commitments and transition to sustainable management. Finally, the agreement should require transparent data documentation and enforcement measures (*9*).

We call on the heads of state of the High Level Panel for a Sustainable Ocean Economy, the Comprehensive and Progressive Agreement for Trans-Pacific Partnership, and the United States–Mexico–Canada Agreement—who have already committed to eliminating harmful subsidies (*10*–*12*)—as well as other trade blocs and individual countries, to declare their support now for an agreement that enshrines these recommendations. WTO members must harness their political mandate to protect the health of the ocean and the well-being of society.

U. Rashid Sumaila[1,2]*, Daniel Skerritt[1], Anna Schuhbauer[1], Sebastian Villasante[3], Andres. M. Cisneros-Montemayor[4], Hussain Sinan[5], Duncan Burnside[1], and 289 additional authors†

[1]Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [2]School of Public Policy and Global Affairs, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [3]Cross-Research in Environmental Technologies, Department of Applied Economics, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain. [4]School of Resource and Environmental Management, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. [5]Marine Affairs Program, Dalhousie University, Halifax, NS B3H 4R2, Canada.
*Corresponding author.
Email: r.sumaila@oceans.ubc.ca
†Full list of authors and affiliations can be found at www.science.org/doi/10.1126/science.abm1680.

REFERENCES AND NOTES

1. D. Pauly *et al.*, *Nature* **418**, 689 (2002).
2. U. R. Sumaila *et al.*, *Mar. Pol.* **109**, 103695 (2019).
3. WTO, Negotiations on fisheries subsidies (2021); www.wto.org/english/tratop_e/rulesneg_e/fish_e/fish_e.htm.
4. D. Tickler *et al.*, *Sci. Adv.* **4**, eaar3279 (2018).
5. E. Sala *et al.*, *Sci. Adv.* **4**, eaat2504 (2018).
6. A. M. Cisneros-Montemayor *et al.*, *Mar. Pol.* **69**, 229 (2016).
7. U. R. Sumaila *et al.*, *ICES J. Mar. Sci.* **65**, 832 (2008).
8. United Nations General Assembly, "Transforming Our World: The 2030 Agenda for Sustainable Development" (2015).
9. M. D. Smith, *Science* **364**, 34 (2019).
10. High Level Panel for a Sustainable Ocean Economy, "Transformations for a sustainable ocean economy: A vision for production, protection and prosperity" (2020); www.oceanpanel.org/ocean-action/files/transformations-sustainable-ocean-economy-eng.pdf.
11. M. A. Khan *et al.*, *J. Econ. Struct.* **7**, 1 (2018).
12. A. M. Villarreal, I. F. Fergusson, "NAFTA and the United States-Mexico-Canada Agreement (USMCA)" (Congressional Research Service Report, 2020).

COMPETING INTERESTS

U.R.S. is an unpaid member of the board of directors of Oceana and science advisory board member of the Pew Charitable Trusts' Global Ocean Legacy project and has received funding from the High Level Panel for a Sustainable Ocean Economy and Oceana. S.V. has received funding from the "EQUALSEA—Transformative adaptation towards Ocean Equity" Project, under the European Horizon 2020 Program.

10.1126/science.abm1680

Alzheimer's drugs: Does reducing amyloid work?

In his Perspective "Treatments for Alzheimer's disease emerge" (6 August, p. 624), D. J. Selkoe asserts that some trials testing potential treatments for Alzheimer's disease have shown "evidence of disease modification." He cites reductions in amyloid plaques (hypothesized to cause cognitive decline) and some modest reductions in cognitive decline shown in four potential drugs that target amyloid. However, hardly any trials have shown an effect, and even the trials with statistically significant results show effects that are too small to be clinically significant or to justify moving forward with the treatments. β-amyloid antibodies can lower amyloid plaques (extracellular aggregated insoluble β-amyloid), but available data show that decreasing amyloid plaques does not in itself lead to reduction in cognitive decline.

PHOTO: XIE MINGFEI/VCG/GETTY IMAGES

The data from six phase 2 or 3 trials of the four medications cited by Selkoe are available in peer-reviewed articles. Four trials were stopped for futility and one trial, lecanemab, was negative (*1*). Only one trial hit its primary end point (for donanemab) (*2*). The negative lecanemab trial did not meet its primary end point at 12 months, despite the potential advantage of a protocol change that created an imbalance in *APOE4* carriers, who experience faster cognitive decline. With only 30% of the treatment cohort composed of *APOE4* carriers compared with 71% of the placebo group, the placebo group would be expected to decline more quickly (*3*, *4*). Selkoe claims that one gantenerumab phase 2 trial reduced amyloid and cognitive decline, but both reported phase 3 gantenerumab trials were stopped for futility and had no significant effects on primary or secondary outcomes (*5*) at 2 years. Selkoe characterizes a small trial of donanemab as "markedly" decreasing amyloid and "significantly" slowing cognitive decline. However, this trial showed only a 3.2-point benefit on a 144-point scale—half the trial team's designated minimally clinically significant effect size (*2*)—and no significant effects on secondary cognitive and functional outcomes (*6*). One aducanumab trial, EMERGE (NCT02484547), showed a 0.39-point (23%) better outcome for the treatment group on the primary Clinical Dementia Rating (CDR-SB) scale outcome at 18 months, but the identical ENGAGE trial showed a 0.03-point (2%) worsening with treatment (*7*).

Selkoe speculates several reasons for failures of past trials but ignores what might be the most obvious: The treatment target (β-amyloid) itself may be wrong. Just as removing smoke does not extinguish a fire, reducing amyloid plaques may not affect the course of Alzheimer's disease. Certainly, trial data do not support any clinical benefit of amyloid plaque reduction (*8*). Neither donanemab nor aducanumab trialists reported an association between amyloid reduction and individual participant clinical outcomes (*2*, *7*). No comparable published results are available for lecanemab or gantenerumab. Furthermore, the amyloid cascade hypothesis proposes β-amyloid aggregation as an early disease trigger, preceding tau phosphorylation and accumulation (*9*). However, despite reducing amyloid plaques, donanemab failed to lower tau and also increased brain atrophy (*10*).

Alzheimer's disease antibody trials represent the definitive test of the amyloid hypothesis of Alzheimer's disease. Objective appraisal of the clinical outcomes data suggests more a failure of hypothesis confirmation than successful translation of this disease model.

Madhav Thambisetty[1]*, Robert Howard[2], M. Maria Glymour[3], Lon S. Schneider[4]
[1]Clinical and Translational Neuroscience Section, Laboratory of Behavioral Neuroscience, National Institute on Aging, Baltimore, MD 21224, USA. [2]Division of Psychiatry, University College London, London W1T 7NF, UK. [3]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94158, USA. [4]Department of Psychiatry and the Behavioral Sciences, Department of Neurology, and the USC Alzheimer Disease Research Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA.
*Corresponding author. Email: thambisettym@mail.nih.gov

REFERENCES AND NOTES

1. C. J. Swanson *et al.*, *Alzheimers Res. Ther.* **13**, 80 (2021).
2. M. A. Mintun *et al.*, *N. Engl. J. Med.* **384**, 1691 (2021).
3. K. R. Murphy *et al.*, *Neuroimage* **78**, 474 (2013).
4. J. Qian *et al.*, *Neurology* **96**, e2414 (2021).
5. S. Ostrowitzki *et al.*, *Alzheimers Res. Ther.* **9**, 95 (2017).
6. A. I. Levey, *N. Engl. J. Med.* **384**, 1762 (2021).
7. U.S. Food and Drug Administration, "Drug Approval Package: Aduhelm (aducanumab-avwa). Statistical Review and Evaluation NDA/BLA #: 761178" (2021); www.accessdata.fda.gov/drugsatfda_docs/nda/2021/761178Orig1s000StatR_Redacted.pdf.
8. S. F. Ackley *et al.*, *BMJ* **372**, n156 (2021).
9. E. S. Musiek, D. M. Holtzman, *Nat. Neurosci.* **18**, 800 (2015).
10. S. Ayton, *Eur. J. Neurol.* **28**, e67 (2021).

COMPETING INTERESTS

L.S.S. has received consulting or advisory fees within the past 3 years from Cortexyme, Cognition Inc., Eisai, Eli Lilly, IBC Ltd., Merck, Neurim Ltd., Roche/Genentech, Samus, Takeda, and UCB and has received grants or contracts from Biogen, Eisai, Eli Lilly, and Novartis related to the subject matter of this correspondence.

10.1126/science.abl8366

Response

Thambisetty *et al.*'s assertion that "trial data do not support any clinical benefit of amyloid plaque reduction" in Alzheimer's disease is not supported by close analysis of all publicly available trial data. Some β-amyloid antibodies but not others have decreased amyloid deposits in brain regions serving cognition, accompanied by slowing of decline on some tests of cognition and activities of daily living. The US Food and Drug Administration's (FDA's) Clinical Pharmacology Review of aducanumab concluded that its reduction of amyloid plaques correlated with slowing of cognitive decline and that this "is consistent across all 6 other programs of β-amyloid antibodies over the past decade" [(*1*), p. 30, see Fig. 5]. The FDA said that the negative aducanumab ENGAGE trial (NCT02477800) was the only exception to this relationship, suggesting that it was a "potential outlier or chance finding" [(*1*), p. 29].

The positive aducanumab EMERGE trial (NCT02484547) showed significantly less (–23%) cognitive decline on 10 mg/kg than placebo (P = 0.0120) at 18 months. Statistically significant treatment effects were also observed for all three ranked secondary clinical end points and the tertiary end point. This was supported by statistically significant dose-dependent reductions of amyloid plaques and of cerebrospinal fluid phosphotau, a marker of tangles that correlate with cognitive decline [(*2*), pp. 1–2]. In post hoc analyses of EMERGE, all chosen comparisons showed a positive effect of aducanumab over placebo (*3*).

In the negative (and non-identical) aducanumab ENGAGE trial (NCT02477800), the FDA noted that patients with the highest drug exposure had benefits on cognition and activities of daily living, like patients in the EMERGE trial with comparable drug exposure [(*2*), p. 3]. The FDA Clinical Pharmacology team did extensive trial simulations and concluded that the probability of the high-dose group in EMERGE being a false positive was very low. They suggested that the negative high-dose group in ENGAGE was likely a chance finding driven in part by the numerous patients who were enrolled before a late protocol amendment allowed *APOE4* carriers to receive the highest dose (10 mg/kg) of aducanumab. The FDA said that the probability of observing the overall aducanumab findings under an assumption that the agent was equal to placebo "was extremely low" [(*2*), p. 5].

Thambisetty *et al.* point out that for donanemab and aducanumab, there was no association between amyloid lowering and individual patient clinical outcomes. But the FDA has stated that randomizing patients into specific dose groups is highly likely to achieve a balance across dose groups (i.e., at group level) of known and unknown prognostic factors that could influence outcomes. If patients are randomized at group level (as they were) but the relationships between end points are assessed at individual level within a dose group, "such a balance (in prognostic factors) can no longer be guaranteed" [(*1*), p. 27]. So, variability among Alzheimer's disease patients may be too great to expect consistent individual correlations with outcomes, but correlations were seen at group level.

The lecanemab Phase 2b trial used Bayesian design and narrowly missed its ambitious 12-month primary outcome, but at 18 months, it showed favorable drug-placebo differences of 27% less decline on one Alzheimer's disease cognitive test and 56% less on another (*4*). However, during enrollment, the regulator prohibited *APOE4* carriers from receiving the highest dose. To

adjust for this, the highest and next-highest dose groups were analyzed together, thereby achieving *APOE4* balance between treatment and placebo groups, and now the first cognitive measure showed 20% less decline on the drug (at a nearly significant $P = 0.053$). Moreover, lecanemab reduced cognitive decline more in *APOE4* carriers than in patients without *APOE4* (*4*), suggesting potentially greater overall benefit if more *APOE4* carriers had been allowed to take the highest dose.

Donanemab in a phase 2 trial robustly lowered amyloid plaques and achieved its primary end point on a composite score of cognition and activities of daily living at 18 months ($P = 0.04$), with nominal positive trends for two secondary cognitive outcomes. Prespecified analyses of Tau–positron emission tomography showed less tau accumulation in frontal and temporal cortices, an important marker of cognitive decline (*5*, *6*).

Thambisetty *et al.* question the clinical meaningfulness of the findings across the trials of these antibodies, but the Alzheimer's disease field has no established quantitative guidelines for what constitutes clinical meaningfulness from a patient and family perspective. In one study of clinically meaningful change, a 1- to 2-point worsening in the Clinical Dementia Rating (CDR-SB) scale was the average change across cognitively normal to moderately severe dementia, but at the level of mild cognitive impairment, a rise in CDR-SB of <1 (.98) was deemed meaningful by judgment of clinicians (not of patients or caregivers) (*7*).

Based on 3 decades of Alzheimer's disease genetics, β-amyloid accumulation is the fire, not the smoke. Clearing amyloid in other diseases (e.g., transthyretin) slows organ failure (*8*), and the above trial results suggest we are starting to see this happen in Alzheimer's disease. Thambisetty *et al.* suggest not moving forward with amyloid-targeted treatments despite growing evidence that these agents can benefit patients, even if only modestly as used so far.

Dennis J. Selkoe
Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.
Email: dselkoe@bwh.harvard.edu

REFERENCES AND NOTES

1. FDA, "FDA clinical pharmacology review" (2021); www.accessdata.fda.gov/drugsatfda_docs/nda/2021/761178Orig1s000ClinPharm_Redacted.pdf.
2. FDA, "Office of Neurology's summary review memorandum" (2021); www.accessdata.fda.gov/drugsatfda_docs/nda/2021/Aducanumab_BLA761178_Dunn_2021_06_07.pdf.
3. Biogen, "Evaluation of aducanumab efficacy in early Alzheimer's disease" (2021), p. 13; https://investors.biogen.com/static-files/74641e1b-cd23-495e-8f29-f4ecac0a1126.
4. C. J. Swanson *et al.*, *Alzheimers Res. Ther.* **13**, 80 (2021).
5. J. S. Sanchez *et al.*, *Sci. Transl. Med.* **13**, eabc0655 (2021).
6. M. A. Busche, B. T. Hyman, *Nat. Neurosci.* **23**, 1183 (2020).
7. J. S. Andrews *et al.*, *Alzheimers Dement.* **5**, 354 (2019).
8. D. Adams, M. Slama, *Curr. Opin. Neurol.* **33**, 553 (2020).

10.1126/science.abm3288

NEXTGEN VOICES: SUBMIT NOW

Postpandemic adjustments

Add your voice to *Science*! In this NextGen Voices survey, a reader asks for your advice. Have you been in this situation or one like it? Do you have any tips that you would like to share? Become a NextGen Voices peer mentor by contributing your thoughts!

> Dear NextGen Voices peer mentors,
> I am so excited that my university is finally allowing us to come back to work in a semi-normal way this year. However, I know that some of my labmates have been exceptionally productive during our time working remotely, publishing papers and finding new grants. Meanwhile, I have been overwhelmed by the stress of being away from my family, who live in a poorer country that has been hit especially hard, while also trying to care for my young children. I haven't published anything, and applying for grants feels impossible. As we move forward, how can I get my already sparse early-career CV (and life!) back on track and compete with all the people who have been so productive during this difficult time?
>
> Sincerely,
> Playing Catchup

To submit, go to www.science.org/nextgen-voices
Deadline for submissions is 5 November. A selection of the best responses will be published in the 7 January issue of *Science*. Submissions should be 150 words or less. Anonymous submissions will not be considered.

NEW!
Open-Design Upright Microscope

- NEW: Optional motorized or fixed XY stage, or motorized translator
- Open-design microscope with motorized focus
- Quickly configurable based on experimental needs
- Optimized *In vivo* and *In vitro* experimentation on one setup
- Uses standard Olympus objectives
- OCC or DIC transmitted light (LED)
- Epi-fluorescent imaging



BOB™

The Sutter BOB – designed to eliminate the conventional microscope frame – is a versatile, open-design upright microscope platform ideal for slice electrophysiology, widefield fluorescent imaging, two-photon imaging, photostimulation and new techniques just being developed!

SUTTER INSTRUMENT
PHONE: +1.415.883.0128
FAX: +1.415.883.0572
EMAIL: INFO@SUTTER.COM
WWW.SUTTER.COM

PRIZE ESSAY



NEUROSCIENCE

The power of hunger

Interoceptive sensing and hunger neurons have a role in the control of behavior

Pain responses are blunted in lab mice that are deprived of nutritive foods such as these high-fat pellets.

PHOTO: AMBER ALHADEFF

By **Amber L. Alhadeff**

Why do we behave differently when we haven't eaten? How do we know when to stop eating? How do the foods that we eat influence our brain activity? Over the past several decades, we have discovered many of the mechanisms that underlie the neural control of food intake and how hunger changes our perception of the world (*1*, *2*). However, the brain constantly receives updates on the physiological condition of the body, and so it is also essential to understand how interoceptive sensory signals that arise from the body influence neural activity in the brain's feeding centers. This component has received comparatively less scholarly attention. My work has sought to connect hunger circuits in the brain with inputs and outputs from the rest of the body, with the goal of discovering how neurons that encode hunger are modulated by what we eat and how their neural activity can change our behavior.

Monell Chemical Senses Center, Philadelphia, PA 19104, USA. Email: aalhadeff@monell.org

HUNGER FILTERS PERCEPTION

How does hunger influence the perception of competing signals? To tackle this question, we first examined how hungry mice respond to different kinds of sensory stimuli. We found that food deprivation causes a profound reduction in behavioral responses to long-term inflammatory pain and itch in mice but leaves intact the ability to respond to acute thermal and mechanical stimuli (*3*, *4*). These findings demonstrate the specificity with which hunger can filter different kinds of noxious sensory information.

We next leveraged our knowledge of the neurons that underlie hunger to investigate the neural circuits that mediate the effect of food deprivation on pain behavior. Hypothalamic agouti-related protein (AgRP)–expressing neurons are critical for food intake control in

eppendorf & Science PRIZE FOR NEURO BIOLOGY

mice and are highly active during food deprivation (*5*, *6*). We therefore hypothesized that activity in this neuron population may underlie the interaction between hunger and pain.

Optogenetic activation of AgRP neurons recapitulated the effects of hunger by selectively inhibiting long-term inflammatory pain caused by injection of a noxious chemical into the mouse paw. To identify the neural circuit that mediates this behavioral interaction, we performed a systematic functional analysis of distinct AgRP neuron subpopulations that project to different brain regions. These experiments revealed a population of only ~300 AgRP neurons that project to the hindbrain parabrachial nucleus (PBN), which eliminates behaviors associated with inflammatory pain. These findings demonstrate the notable specificity through which a hypothalamic-to-hindbrain circuit suppresses competing survival behaviors during hunger.

We further refined our understanding of this circuit by demonstrating that neuropeptide Y (NPY) released from AgRP neurons inhibits pain. We also restored pain responses in hungry mice by blocking NPY Y1 receptor signaling in the PBN, highlighting the physiological relevance of this analgesic mechanism (*3*).

These findings place hunger circuits in a neural hierarchy that filters and processes sensory information to prioritize behavior. We are optimistic about the prospect of using hunger circuits as a starting point for the development of safe and effective pain therapeutics. Determining the precise mechanisms through which central hunger circuits inhibit ascending sensory information has the potential to revolutionize our ability to treat chronic pain.

GUT-BRAIN SIGNALING AND FOOD INTAKE

Hunger circuits can clearly have profound effects on an organism's responses to sensory stimuli, but how do sensory stimuli affect neural activity in hunger circuits? We next monitored AgRP neuron activity in awake, behaving mice to determine how sensory information in the gastrointestinal tract affects activity in hypothalamic neurons.

There are multiple qualities of food that influence neural activity and behavior (*7*–*9*), including external sensory cues (e.g., visual, olfactory, and taste) and the nutritive content of food. We sought to determine the aspects of food intake that are most important for the regulation of AgRP neurons. By examining the effects of caloric and noncaloric foods on neural activity, we showed that calories are required for sustained reductions in AgRP neuron activity (*10*). Direct nutrient infusion into the stomach was sufficient to suppress AgRP neuron activity, further suggesting a key role for calories in mediating these changes. Indeed, we showed that calories train sensory cues to inhibit AgRP neuron activity in a single trial (*10*).

What are the gut-brain pathways through which nutritive signals ultimately inhibit AgRP neurons? We performed a series of experiments to manipulate different gut-brain pathways while monitoring in vivo AgRP neuron activity and found that whereas fat engages vagal signaling to inhibit AgRP neuron activity, the vagus nerve is not required for the inhibition of AgRP neuron activity by glucose (*11*, *12*). Rather, sugar engages spinal afferent signaling pathways that communicate with hypothalamic hunger neurons (*12*). These unexpected findings highlight the role of an understudied gut-brain pathway—spinal afferents—in nutrient sensing and provide insight into how different nutrients affect our brains and therefore our behavior.

Determining the pathways through which gut signals influence key feeding centers in the brain has important implications for diseases related to weight control. For example, a better understanding of how gut-brain pathways are engaged by particular nutrients may provide insight into why some foods, such as those that are high in fat and sugar, are more rewarding than others. Because AgRP neurons drive food intake by transmitting a negative valence signal (*13*), understanding the pathways for AgRP neuron inhibition may help ease the negative feelings associated with hunger, and therefore improve weight-loss efforts.

BRIDGING THE BRAIN-PERIPHERY DIVIDE

Research has begun to unravel the mechanisms through which the brain interacts with the rest of the body, and vice versa, but our understanding of this complex and multifaceted communication is incomplete. Moving forward, research in my laboratory will focus on understanding these interactions. Ultimately, disentangling these mechanisms will aid in the development of targeted therapies for a broad range of human diseases that involve pain and weight control. ■



WINNER

Amber L. Alhadeff

Amber L. Alhadeff received her undergraduate degree and PhD from the University of Pennsylvania. After also completing her postdoctoral fellowship there, Amber started her laboratory at the Monell Chemical Senses Center and the Department of Neuroscience at the University of Pennsylvania in 2020. Her research investigates gut-brain signaling and its contributions to feeding and other motivated behaviors.



FINALIST

Justin Rustenhoven

Justin Rustenhoven received an undergraduate degree and PhD in pharmacology from the University of Auckland in New Zealand. He performed his postdoctoral training in the laboratory of Professor Jonathan Kipnis at Washington University in St. Louis, where he continues to work as a research fellow. His research explores the mechanisms underlying immune surveillance of the central nervous system from the brain's borders.
science.org/doi/10.1126/science.abl7122



FINALIST

Andreas J. Keller

Andreas Keller received his undergraduate and graduate degrees in physics from ETH Zürich. His PhD work, completed under Dr. Kevan Martin, focused on rapid network state transitions in the visual cortex. He joined the laboratory of Dr. Massimo Scanziani at the University of San Francisco for his postdoctoral training. Andreas started his laboratory at the Institute of Molecular and Clinical Ophthalmology Basel in 2021, where his research focuses on mechanisms of cortical plasticity in feedforward and feedback circuits. **science.org/doi/10.1126/science.abl7124**

REFERENCES AND NOTES

1. S. M. Sternson, A. K. Eiselt, *Annu. Rev. Physiol.* **79**, 401 (2017).
2. A. K. Sutton, M. J. Krashes, *Trends Endocrinol. Metab.* **31**, 495 (2020).
3. A. L. Alhadeff *et al.*, *Cell* **173**, 140 (2018).
4. A. L. Alhadeff, O. Park, E. Hernandez, J. N. Betley, *Neuroscience* **450**, 126 (2020).
5. S. Luquet, F. A. Perez, T. S. Hnasko, R. D. Palmiter, *Science* **310**, 683 (2005).
6. K. A. Takahashi, R. D. Cone, *Endocrinology* **146**, 1043 (2005).
7. J. N. Betley *et al.*, *Nature* **521**, 180 (2015).
8. Y. Mandelblat-Cerf *et al.*, *eLife* **4**, e07122 (2015).
9. Y. Chen, Y. C. Lin, T. W. Kuo, Z. A. Knight, *Cell* **160**, 829 (2015).
10. Z. Su, A. L. Alhadeff, J. N. Betley, *Cell Rep.* **21**, 2724 (2017).
11. A. L. Alhadeff *et al.*, *Neuron* **103**, 891 (2019).
12. N. Goldstein *et al.*, *Cell Metab.* **33**, 676 (2021).
13. J. N. Betley *et al.*, *Nature* **521**, 180 (2015).

10.1126/science.abl7121

PHOTOS: (TOP TO BOTTOM) HAPPY HOUR HEADSHOTS; MIKE NICHOLS/MEDICAL PORTRAIT STUDIO; FLUSSREIF

PRIZE ESSAY



FINALIST
Justin Rustenhoven
Justin Rustenhoven received an undergraduate degree and PhD in pharmacology from the University of Auckland in New Zealand. He performed his postdoctoral training in the laboratory of Professor Jonathan Kipnis at Washington University in St. Louis, where he continues to work as a research fellow. His research explores the mechanisms underlying immune surveillance of the central nervous system from the brain's borders.
science.org/doi/10.1126/science.abl7122

NEUROSCIENCE

A privileged brain

Local neuroimmune collaboration safeguards the central nervous system

By **Justin Rustenhoven**

Our immune system and central nervous system (CNS) continuously surveil the environment and make on-demand adjustments to maintain homeostasis. Yet despite their commonalities, these systems were originally believed to be entirely separate from one another.

Unlike peripheral tissues, the CNS lacks a parenchymal lymphatic network, and the blood–brain barrier and blood–cerebrospinal fluid (CSF) barrier restrict homeostatic immune cell entrance (*1*). This knowledge fueled a long-held dogma of CNS "immune privilege," which suggested an inability of the brain and spinal cord to mount a functional immune response to local insults (*2*, *3*). But why would our brain and spinal cord, arguably our most complex and delicate organs, lack such a fundamental arrangement to detect dysfunction? We posited the complex nature of the CNS necessitates its own specialized immunological setup to detect and respond to perturbations accordingly (*4*).

eppendorf & Science PRIZE FOR NEURO BIOLOGY

SURVEILLING CNS BORDERS

The flushing of brain interstitial fluid into the surrounding CSF revealed a potential pathway facilitating CNS immune surveillance (*5*). This clearance mechanism shuttles the soluble immunogenic constituents of the brain, including microbial pathogens, tumor neoantigens, and even self-antigens, out of the CNS and enables access to peripheral tissues. We therefore reasoned that the detection of these soluble antigens could allow the immune system to continuously probe for potential CNS dysfunction and explored the locations and mechanisms mediating such monitoring.

Our research suggests that neuroimmune surveillance is concentrated at the edges of the CNS, in brain-border tissues called the dural meninges (*6*). Instead of CSF flowing into venous blood (*7*), our imaging revealed CNS-derived antigens in the CSF rapidly accumulated in the dural meninges of the mouse, specifically surrounding large vascular structures called the dural sinuses (*6*). Demonstrating evolutionary conserved fluid dynamics, this pathway was also visualized by magnetic resonance imaging in human patients (*8*), while liquid chromatography–mass spectrometry revealed CNS-enriched antigens surrounding human dural sinuses (*6*). Critically, the dural sinuses also contained highly organized hubs of antigen-presenting cells capable of capturing brain-derived antigens and presenting them on their surface for recognition by another specialized immune cell, the T cell.

All components necessary for CNS antigen presentation are concentrated at the dural sinuses, but can adaptive immune cells surveil this tissue? We generated a single-cell RNA sequencing (scRNA-seq) meningeal atlas and found that the dural sinuses contain a stromal niche perfectly primed for T cell extravasation via specialized endothelium and stromal-derived chemokines (*6*). Using in vivo imaging and transfer of antigen-specific T cells, we also confirmed that, unlike the CNS vasculature, T cells constantly infiltrate this site and, upon recognition of CSF-derived antigens, initiate an effector response designed to restore brain homeostasis.

These findings reshaped our understanding of neuroimmune surveillance, highlighting the dural meninges as a critical site for the recognition of CNS perturbations. The question then became: What do the meninges do with this information?

BESPOKE IMMUNE CELLS FOR THE CNS

There is a large pool of bone marrow situated on the periphery of the brain, in the skull overlying the dural sinuses. Here, active hematopoiesis—the formation of new blood cells—occurs via resident lymphoid and myeloid progenitors (*9*). In mice and humans, this bone marrow niche is directly connected to the underlying dura matter by way of ossified channels, providing direct and uninterrupted passage between the bone marrow of the skull and the underlying dural meninges (*10*, *11*). Parabiotic pairing, skull transplantation models, and selective irradiation of discrete bone marrow niches were used in mice to demonstrate that these channels are con-

Department of Pathology and Immunology, Washington University in St. Louis, St. Louis, MO 63110, USA. Email: rustenhoven@wustl.edu

PHOTO: MIKE NICHOLS/MEDICAL PORTRAIT STUDIO

duits for homeostatic myeloid and lymphoid cell trafficking from the skull and vertebral bone marrow (*12*).

These vascular highways allow B cells, neutrophils, and monocytes to make their way into the underlying dural tissue and provide substantial contributions to its immune repertoire, wholly independent of a blood route (*12*, *13*). Such a pathway allows for the delivery of a particular subset of immune cells directly to CNS border tissues. We hypothesized that the neuroimmune surveillance afforded to the dural sinuses ideally positions them to interpret CNS dysfunction and relay this to adjacent bone marrow, which can then deploy immune cells in response to a variety of insults (*6*).

After neuroinflammatory injury induced by experimental autoimmune encephalomyelitis (EAE), a mouse model of multiple sclerosis, elevated numbers of CNS bone marrow–derived cells were present in the dural meninges and were no longer restricted to the CNS borders (*12*). Monocytes infiltrating the CNS were largely of local bone marrow origin and frequently associated with bone marrow channels. Similar patterns were observed after optic nerve and spinal cord injuries, demonstrating a conserved ability of CNS-associated bone marrow–derived cells to infiltrate CNS tissue in the wake of varied pathologies (*12*).

Not only does this local bone marrow function as an immunological reservoir, but the cells originating from this niche are phenotypically distinct. Through scRNA-seq of CNS-infiltrating cells following EAE and spinal cord injury, we detected elevated levels of pathogenic proinflammatory cytokines and chemokines—which are implicated in worsened pathology (*14*)—in blood-derived cells compared with their bone marrow–derived counterparts (*12*). Cells originating from local bone marrow likely play distinct roles in CNS injury, shaped by their specific ontogeny and interactions with stromal niches during transit through the dura (*6*, *15*).

COLLABORATIVE CARE

The complexities of neuroimmune interactions are continuously being unraveled. Our findings have uncovered an immunological orchestration that permits the CNS to maintain its barriers and enable tight regulation of cerebral function, while still allowing antigenic sampling to recognize CNS perturbations (*6*), and for the subsequent mobilization of tailor-made immune cells from local bone marrow reservoirs (*12*). Through this clever compromise, the CNS has pushed its immunity to its borders, allowing uninterrupted neuronal function, without completely sacrificing immunological protection. ■

REFERENCES AND NOTES

1. K. Alves de Lima, J. Rustenhoven, J. Kipnis, *Annu. Rev. Immunol.* **38**, 597 (2020).
2. M. J. Carson, J. M. Doose, B. Melchior, C. D. Schmid, C. C. Ploix, *Immunol. Rev.* **213**, 48 (2006).
3. P. B. Medawar, *Br. J. Exp. Pathol.* **29**, 58 (1948).
4. J. Rustenhoven, J. Kipnis, *Science* **366**, 1448 (2019).
5. J. J. Iliff *et al.*, *Sci. Transl. Med.* **4**, 147ra111 (2012).
6. J. Rustenhoven *et al.*, *Cell* **184**, 1000 (2021).
7. S. T. Proulx, *Cell. Mol. Life Sci.* **78**, 2429 (2021).
8. G. Ringstad, P. K. Eide, *Nat. Commun.* **11**, 354 (2020).
9. C. Christodoulou *et al.*, *Nature* **578**, 278 (2020).
10. F. Herisson *et al.*, *Nat. Neurosci.* **21**, 1209 (2018).
11. R. Cai *et al.*, *Nat. Neurosci.* **22**, 317 (2019).
12. A. Cugurra *et al.*, *Science* **373**, eabf7844 (2021).
13. S. Brioschi *et al.*, *Science* **373**, eabf9277 (2021).
14. A. Giladi *et al.*, *Nat. Immunol.* **21**, 525 (2020).
15. I. Ballesteros *et al.*, *Cell* **183**, 1282 (2020).

10.1126/science.abl7122

PRIZE ESSAY



FINALIST
Andreas J. Keller
Andreas Keller received his undergraduate and graduate degrees in physics from ETH Zürich. His PhD work, completed under Dr. Kevan Martin, focused on rapid network state transitions in the visual cortex. He joined the laboratory of Dr. Massimo Scanziani at the University of San Francisco for his postdoctoral training. Andreas started his laboratory at the Institute of Molecular and Clinical Ophthalmology Basel in 2021, where his research focuses on mechanisms of cortical plasticity in feedforward and feedback circuits. **science.org/doi/10.1126/science.abl7124**

NEUROSCIENCE

Hidden in plain sight

Context controls the activity of sensory neurons

By **Andreas J. Keller**

Where is Waldo? Despite his striking red-and-white–striped shirt, Waldo (frequently known as Wally outside the United States) of the popular children's puzzle-book series "Where's Waldo?" is consistently difficult to find in the crowded illustrations in which he is embedded. Readers must search a given scene meticulously to find him and are often fooled by patterns surrounding the elusive protagonist that resemble those of his shirt. Why is this seemingly simple task so challenging for our visual systems?

Our perception of a visual stimulus is strongly influenced by the scene surrounding the stimulus—the visual context. Depending on its context, the same visual stimulus may be perceived as more or less salient, allowing it to stand out or meld into the rest of the visual scene. This contextual modulation is what leads to a reduction in our visual system's response to Waldo's shirt. If he appeared alone, or was at least the only one wearing a red-and-white–striped shirt, he would be conspicuous and spotted effortlessly.

SEEING THE STIMULUS AND THE SURROUND

Context gives meaning to sensations. Accordingly, sensory processing of a stimulus is shaped by the context in which the stimulus is embedded. My colleagues and I set out to understand the neural circuits in the mouse visual cortex responsible for contextual modulation of sensory stimuli.

Every sensory neuron in the visual system receives input from one region in space and responds to stimuli placed in that region. Adjacent neurons in the visual cortex receive input from surrounding regions in space. The responses of excitatory neurons are suppressed when stimulus and surround are similar but are facilitated when they differ. Somatostatin-expressing (SOM) inhibitory neurons integrate information from the surrounding regions and provide surround suppression to other neurons (*1*). Could these inhibitory neurons also account for surround facilitation when stimulus and surround differ?

Using optical recordings in the visual cortex of awake mice, we found that SOM neurons are active when stimulus and surround are similar but not when they differ. Conversely, inhibitory neurons expressing vasoactive intestinal peptide (VIP)—the main inhibitor of SOM neurons—are particularly active when stimulus and surround differ (*2*).

COORDINATING CONTEXTUALIZATION

On the basis of these results and known connectivity motifs between VIP, SOM, and excitatory neurons (*3*–*5*), we hypothesized a simple mechanism to explain contextual modulation in excitatory neurons: When stimulus and surround are similar, SOM neurons are active and suppress excitatory neurons. When stimuli and surround differ, VIP inhibitory neurons are active, suppress SOM neurons, and thereby relieve excitatory neurons from suppression. Using computational modeling, optical recordings, and optogenetic perturbations, we showed that VIP neurons orchestrate the amount of suppression that SOM neurons provide to other neurons (*2*). VIP neurons may coordinate cortical circuits in adjusting the saliency of a stimulus in a visual scene.

MAKING THE PICTURE WHOLE

Current dogma holds that neurons respond to stimuli and are modulated by the visual context. Context, however, can play more than a supporting role. When a neuron's visual information is missing—for example, because of obstruction—its activity can be driven by context alone (*6*–*12*), which allows our sensory neurons to fill in the gaps. Likely relying upon past visual experience, missing information is inferred from the context.

The visual system comprises several stages of processing. Information flows from the retina to the primary visual cortex, from where it is then sent to higher visual areas. Information also flows in the opposite direction through feedback projections from higher visual areas to earlier stages of processing. It is this reverse flow of information that is responsible for filling the gaps when visual information is missing (*12*).

Institute of Molecular and Clinical Ophthalmology Basel (IOB), 4057 Basel, Switzerland. Email: andreas.keller@iob.ch

PHOTO: FLUSSREIF

FEEDBACK FILLS IN THE GAPS

What is the mechanistic basis for this fundamental perceptual phenomenon? Using optical recordings in awake mice, we found that neurons in the primary visual cortex—but not in earlier stages of processing—respond to context alone. Contextual drive must come from neurons that respond to a stimulus in a surrounding region. These neurons could be in the primary visual cortex, in higher visual areas, or even in an earlier stage of processing. However, each of these scenarios makes different predictions related to the time required for a signal to travel from the retina to the neurons in the primary visual cortex.

If contextual drive is inherited from surrounding neurons in an earlier stage of processing, we would expect transmission to take as much time as a signal coding for a stimulus in the center (the additional cable length connecting the surrounding neurons is negligible). Otherwise, we would expect there to be at least one additional neuron in the processing chain, which would generate a measurable delay. We found a delay between the signal from a centered stimulus and one from the context, making it unlikely that contextual drive directly originates in earlier stages of processing.

To test whether responses to the visual context are driven by activity in higher visual areas, we developed a method that allowed us to silence neurons in higher visual areas while recording the visual responses of neurons in the primary visual cortex. By silencing higher visual areas, we virtually abolished the contextual drive. Using optical recordings from axons originating in higher visual areas and projecting to the primary visual cortex, we found that these feedback projections carry the information necessary for contextual drive—this reverse flow of information likely filling in missing information (*12*). In other words, when there is nothing to see, we see what we expect to see: Context dominates when Waldo is hidden in plain sight. ■

REFERENCES AND NOTES

1. H. Adesnik, W. Bruns, H. Taniguchi, Z. J. Huang, M. Scanziani, *Nature* **490**, 226 (2012).
2. A. J. Keller *et al*., *Neuron* **108**, 1181 (2020).
3. C. K. Pfeffer, M. Xue, M. He, Z. J. Huang, M. Scanziani, *Nat. Neurosci*. **16**, 1068 (2013).
4. X. Jiang *et al*., *Science* **350**, aac9462 (2015).
5. H.-J. Pi *et al*., *Nature* **503**, 521 (2013).
6. R. von der Heydt, E. Peterhans, G. Baumgartner, *Science* **224**, 1260 (1984).
7. Z.-M. Shen, W.-F. Xu, C.-Y. Li, *J. Physiol.* **583**, 581 (2007).
8. M. Fiorani Júnior, M. G. Rosa, R. Gattass, C. E. Rocha-Miranda, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8547 (1992).
9. A. F. Rossi, R. Desimone, L. G. Ungerleider, *J. Neurosci*. **21**, 1698 (2001).
10. H. E. Jones, K. L. Grieve, W. Wang, A. M. Sillito, *J. Neurophysiol*. **86**, 2011 (2001).
11. U. H. Schnabel *et al*., *Sci. Rep*. **8**, 17800 (2018).
12. A. J. Keller, M. M. Roth, M. Scanziani, *Nature* **582**, 545 (2020).

10.1126/science.abl7124

AAAS video series focuses on combating misinformation

Vaccines, AI, and sexually transmitted infections among first topics

By **Becky Ham**

Since its earliest days in 2020, the COVID-19 pandemic has surged across the globe with a twin pandemic of misinformation about the disease's causes, origins, treatments, and prevention. The rise in scientific misinformation was on the mind of American Association for the Advancement of Science (AAAS) videographer and multimedia editor Neil Orman when he created a new video series showing how AAAS-affiliated researchers are pushing back against this trend.

"AAAS Voices: Countering Science Misinformation" (www.aaas.org/news/misinformation) interviews researchers who have made combating misinformation a significant part of their careers. In five videos launched this spring and fall, scientists discussed some of the most prominent myths and false claims in their fields, while providing trusted resources and strategies to deliver those resources.

The series has discussed misinformation surrounding vaccines with Maria Elena Bottazzi, associate dean of the National School of Tropical Medicine at Baylor College of Medicine; climate change with Astrid Caldas, a senior climate scientist at the Union of Concerned Scientists; HIV with Collene Lawhorn, a program officer in the Division of AIDS Research at the National Institute of Mental Health; artificial intelligence with Anita Nikolich, director of research and technology innovation at the University of Illinois; and sexually transmitted infections (STIs) with Ina Park, a professor of family and community medicine and obstetrics and gynecology at the University of California, San Francisco. All participants are former AAAS Leshner Leadership Institute for Public Engagement fellows or AAAS Science and Technology Policy (STPF) fellows.



Peter Hotez (l) and Maria Elena Bottazzi regularly share vaccine information with the public.

PHOTO: BAYLOR COLLEGE OF MEDICINE

"AAAS.org was moving in the direction of creating more original content to further position the organization as a thought leader in STEM. The Misinformation Series fit perfectly into this new approach," AAAS Director fo Communications Risa Dixon said.

Orman fleshed out the idea with previous supervisor Joel Goldberg, then brought the idea for a series to Dixon. Together with AAAS senior writer Andrea Korte they began brainstorming topics. The goal, Orman said, was not only to highlight areas of misinformation but also to share concrete ways scientists are tackling the problems.

"The series is positive in tone, focusing on constructive efforts to provide accurate science information," Orman said. "It allows them to talk frankly about productive things they're doing and suggestions they have."

In her video, Park talked about the myths surrounding STIs. A former Leshner Leadership fellow, she offered a history of the stigma attached to STIs and discussed the strategies she's used to push past that stigma to promote healthier discussions, such as encouraging health care providers to share their own stories. "Those personal narratives go a long way in countering things that folks read online," she said, "and when they come from a trusted source like yourself, they are more likely to sway patients."

The Leshner Leadership fellows and STPF have been a rich resource of experts for the videos. "When we asked for fellows who were working to address scientific misinformation, responses were many, swift, and enthusiastic," said Kat Song, the communications and digital strategy director for the STPF. "Mass communication is intrinsic to the strength of democracy, and lots of STPF fellows are both professionally and personally committed to advancing good public information."

Bottazzi said she "jumped at the opportunity" to participate in the series, "not only because I feel it's my responsibility to convey my story and my message, but also because AAAS has a wonderful reach with the science community and general public."

Her time as a Leshner Leadership fellow helped her become more comfortable with speaking about vaccines with a wider variety of audiences, said Bottazzi, who was born in Italy and grew up in Honduras. She is interested in adapting some of those lessons "to the Latinx communities, to make sure that it fits the people that I speak to, with a focus on empowering and decolonizing some of this communication, so that peers and colleagues and younger generations in my countries of origin get involved in doing these kinds of communication."

Each video shares some of the personal background of the researcher, which was a deliberate storytelling choice by Orman. "I want to know about the credentials they bring and why they wanted to do this in the first place," he said. "I think it can legitimize these voices as people who have depth and experience in the field."

Nikolich said speaking about her own career journey in the video was unusual for her but hopes it will help her message resonate with audiences. She received "across the board really good feedback" about the video from her colleagues in academia and the AI Village at DEF CON, a community of hackers and data scientists who share information on the use and abuse of artificial intelligence in security and privacy.

In early October, Nikolich received a grant from the National Science Foundation (NSF) to build a game to teach senior citizens and middle school students about AI misinformation, partnering with some of the developers behind the wildly popular online game World of Warcraft. "I know for a fact that NSF had watched the video," she said.

One of the lessons Nikolich has learned about combating misinformation is that "it's really hard in an immediate conversation to change people's minds because people are flooded with misinformation. All you can do as a scientist is stick to the facts, try not to be emotional and understand that people are swayed by social media."

Orman would like to continue the series with a new group of experts and topics next year. "I've been very grateful to the participants and the AAAS colleagues who suggested these great voices," he said. "It drives home the incredible programs we have and the incredible people who pass through them."



THE MANY BENEFITS OF HEALTHY SLEEP

By **Peter Stern**

Every night we lie down, close our eyes, and lose consciousness for several hours. We are not only inactive but also vulnerable.

Modern society has long underrated sleep. Recently, we have seen a shift toward the appreciation of a good night's sleep. The myth of the strong, successful individual who can get by with minimal time in bed or who skimps on sleep to get more done has by now been thoroughly debunked. Sleep is an important and non-negotiable bodily function that is essential for human health. It benefits our brains and mental well-being, and even other entities such as our cardiovascular and immune systems.

Because sleep is so common in humans and in the animal world, we take it for granted and do not ask the obvious question: What might be its physiological purpose? The simplistic answer—that our bodies and nervous systems need to rest—can be dismissed. We may not be aware of it, but our brains are highly active while we sleep. Over the past decades, we have learned much about the genes, molecules, cells, and circuits involved in sleep and its regulation. Yet we still do not really know why we sleep.

This special issue summarizes recent insights into the basic mechanisms underlying sleep and the many functions our brains perform during this time. The issue also focuses on sleep's restorative and health-promoting aspects for individuals and for society more generally. It is time to acknowledge and enjoy what Shakespeare called the "honey-heavy dew of slumber."

INSIDE

NEWS

Divided we sleep p. 552

REVIEWS

The inescapable drive to sleep: Overlapping mechanisms of sleep and sedation p. 556

Brain neural patterns and the memory function of sleep p. 560

The interconnected causes and consequences of sleep in the brain p. 564

The translational neuroscience of sleep: A contextual framework p. 568

RELATED ITEMS

NEWS p. 526

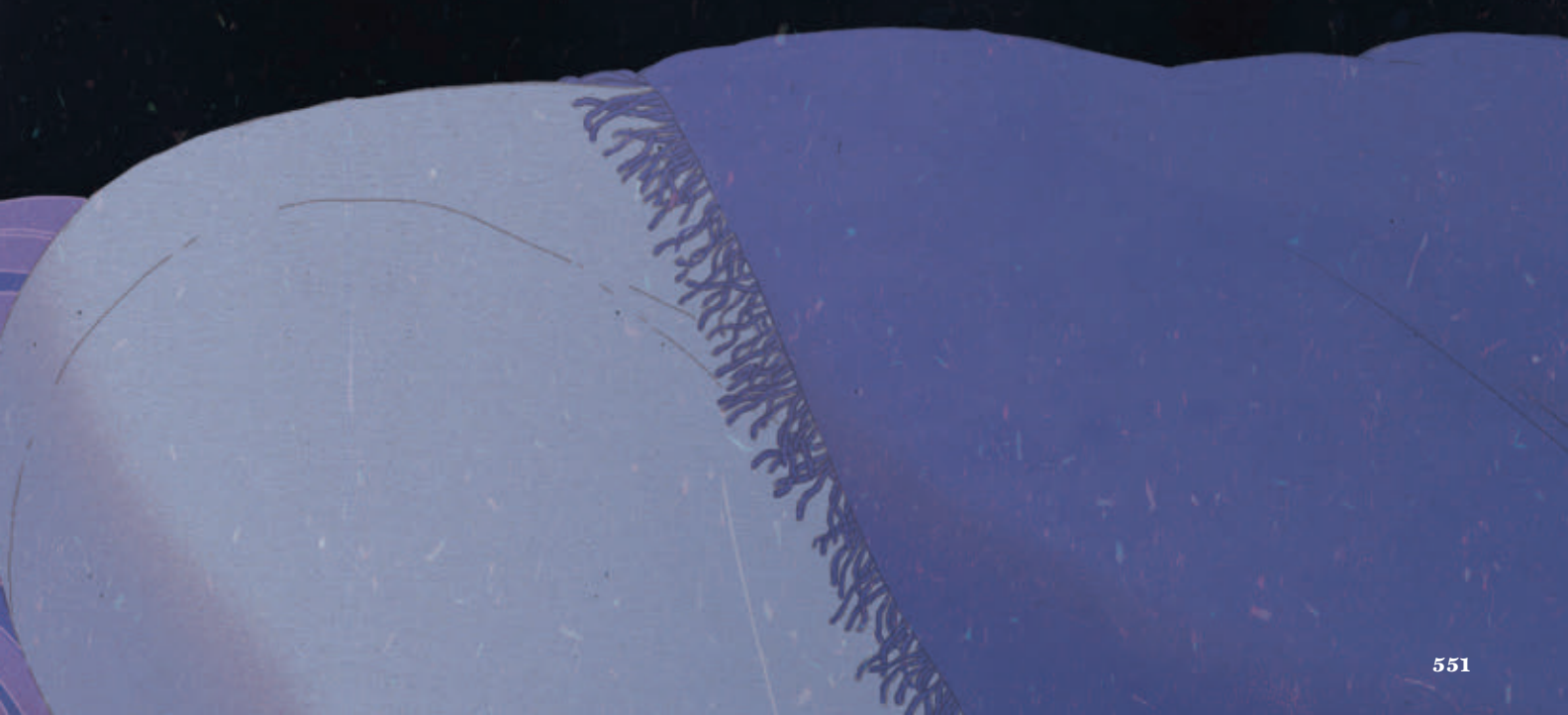
POLICY FORUM p. 530

PODCAST

VIDEO

One of the many functions performed by our brains during sleep is the consolidation of memories. Recent episodes in our daily lives are processed for long-term storage.

ILLUSTRATION: DION MBD





DIVIDED WE SLEEP

Poor sleep disproportionately undermines the health of communities of color. Researchers want to figure out why—and find solutions

By **Rodrigo Pérez Ortega**

PHOTO: PROSTOCK-STUDIO/ALAMY STOCK PHOTO

On a Sunday morning earlier this month, Girardin Jean-Louis attended the regular mass service at Trinity Baptist Church in North Miami, Florida. He wasn't just there to worship, though. After the hourlong service, he stepped to the lectern, microphone in hand, and spoke to the congregation—most of them Black—about the importance of good sleep. "I was using the Bible, talking about how important sleep used to be for all of the kings and queens," he says. "They always had a sage [who] would interpret the dreams, so I use this to show dreaming is very important."

A sleep researcher at the University of Miami, Jean-Louis splits his time between the lab, where he studies the disproportionate toll of sleep deficiency on communities of color, and places at the center of many Black communities: churches, barbershops, beauty salons. There, with the help of pastors, barbers, and other influential figures, he preaches the importance of sleep. Being a Baptist helps, too, he says. "I know the language, I know the lingo."

He is trying to address a historically neglected aspect of racial inequity. Far more Black and Hispanic people than white people report routinely getting less than 6 hours of sleep, well short of the recommended 7 to 9 hours for adults. Studies have shown nonwhite people take longer to fall asleep, wake up more during the night, and spend less time in deep sleep—a stage of sleep important for strengthening the immune system, forming memories, and restoring the health of cells. Sleep disorders, such as obstructive sleep apnea, are up to twice as common in people of color. All of this increases the risk of chronic diseases aggravated by poor sleep. "Sleep is a privilege," says Dayna Johnson, a sleep epidemiologist at Emory University. "If we can target sleep, we might be able to reduce the burden of all types of diseases among racial minorities."

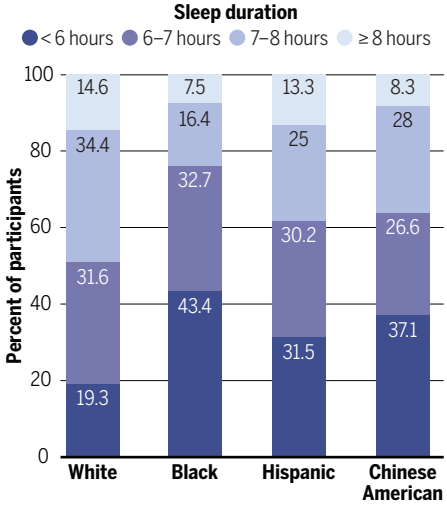
Researchers like Johnson and Jean-Louis are trying to understand what causes these disparities and how scientists, physicians, and policymakers might implement solutions—and begin to right what they see as an underappreciated source of social and health injustice. "Sleep deficiency is literally undermining the fabric of society, particularly in minority and low socioeconomic populations, because they bear the disproportionate burden of the associated risk," says neurobiologist Marishka Brown, director of the National Center on Sleep Disorder Research at the National Heart, Lung, and Blood Institute. "Sleep is a modifiable source of inequity."

Researchers are starting to understand the social and environmental factors that contribute to poor sleep—and increase the risk of health problems—in communities of color.

JEAN-LOUIS FIRST became interested in sleep disparities when he was a postdoctoral researcher in psychiatry at the University of California, San Diego, in the late 1990s. In a survey study, he and colleagues found nonwhite men in San Diego slept about 1 hour less per night on average than white men. Jean-Louis recalls wondering why at the time: "Is it a function of health care access? Is it a function of lifestyle factors?"

Another inequity

A 2015 study of 2230 people in six U.S. communities found people of color were much more likely than white people to sleep less than 6 hours per night on average.



Some evidence points to biological factors—such as differences in upper airway anatomy that might make people of Asian descent more prone to sleep apnea, in which soft throat tissue obstructs the airway and causes waking. But most researchers think socioeconomic and environmental factors are the main cause of these disparities. "I want to move attention away from those genetic explanations because what that means is that it's that person's individual responsibility to fix that issue, versus shining a light on the structural and social factors," says Carmela Alcántara, a clinical psychologist at Columbia University.

In recent decades, researchers have identified some of those social and environmental factors. For example, more Black and Hispanic people work nontraditional hours, including night shifts, which makes it difficult for them to get a good night's sleep. A 2010 study of 340 workers at an extended-care facility in Massachusetts, for example, found Black and Hispanic employees were twice as likely as white ones to work the night shift and slept up to 1 hour less per day on average than white people in general.

Although shift work affects sleep in all ethnic groups, Hispanic people, especially those who are unemployed, may experience a more specific sleep disrupter. Alcántara's research has found acculturation stress—distress that results from adapting and integrating into a new culture—increases the severity of insomnia symptoms. "There's something unique for this population about this specific sociocultural stressor." In addition, discrimination related to their ethnicity and status as immigrants also affects sleep in Hispanic people, she says.

Indeed, more than a dozen studies have identified racial discrimination as a contributor to sleep disparities. In one published last year, Johnson and colleagues asked 1458 people in Detroit who met clinical criteria for insomnia how often they have been treated badly or unfairly because of their race or ethnicity. Black people reported more discrimination and more severe insomnia symptoms than white people, and a statistical analysis determined discrimination accounted for 60% of their insomnia severity.

Environmental factors also cut into sleep. A 2020 study that combined data from the U.S. census with an atlas of light pollution derived from satellite data found Black, Hispanic, and other people of color tend to reside in brighter areas, where they are exposed to approximately twice as much ambient light at night as white people. Exposure to artificial light from the street and commercial buildings has been found to suppress melatonin—a hormone that helps initiate sleep. That causes people to fall asleep later at night, resulting in poorer sleep overall.

Black, Hispanic, and Asian people in the United States are also exposed to disproportionately high levels of particulate air pollution. Exposure to this type of pollution can cause inflammation of the nose and throat, and some evidence suggests chronic exposure can worsen sleep apnea and increase daytime sleepiness. And then there is nighttime noise, which a 2017 study at hundreds of sites across the United States found is higher in neighborhoods with a higher proportion of Black residents.

"Our ZIP code is as—if not more—important than our genetic code," Johnson says.

"SLEEP IS THE MOST primordial human activity," says Azizi Seixas, a precision medi-

CREDITS: (GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) CHEN *ET AL.*, *SLEEP*, **38**, 6, 2015

cine researcher at New York University's (NYU's) Langone Health. "We can't do without it." Besides leading to drowsiness and impaired cognition, lack of sleep also takes a toll on the body.

As early as the 1980s, studies linked sleep apnea and shift work with hypertension and coronary heart disease. When we sleep, our blood pressure goes down, Johnson explains. "We give our blood vessels a break, essentially," she says. But if sleep is interrupted or shortened, the body maintains a higher blood pressure at night—what physicians call nocturnal hypertension—which raises the risk of heart disease and stroke. Many studies have shown Black shift workers have higher chances of developing hypertension, compared with white shift workers.

Impacts of poor sleep

Sleeping 6 hours or less per night has detrimental effects throughout the body. Light and noise pollution, discrimination, stress, socioeconomic status, and shift work are some of the factors that affect sleep in communities of color.

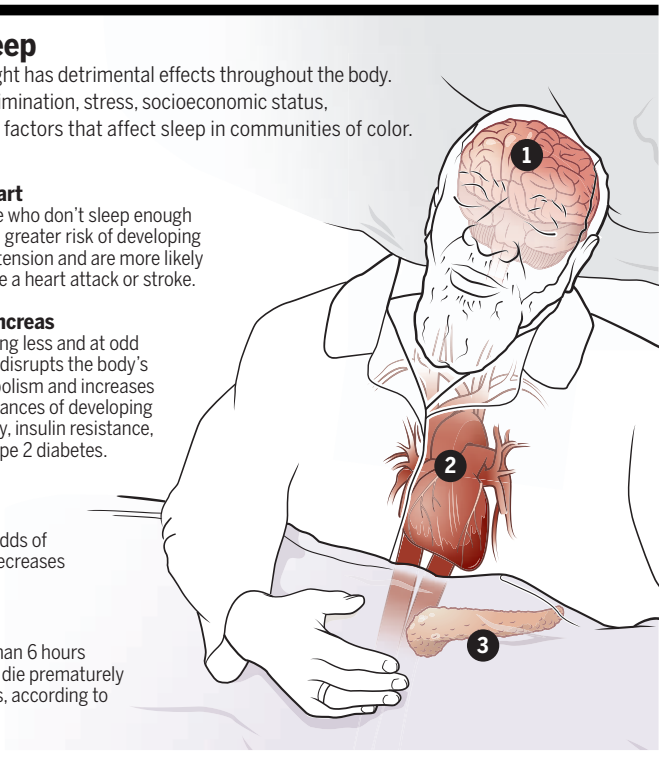
1 Brain
Poor sleep decreases alertness, which can contribute to accidents at work and on the road. Bad sleep also impairs cognition and worsens depression and anxiety symptoms. Chronic sleep problems may predispose people to dementia—particularly Alzheimer's disease.

2 Heart
People who don't sleep enough have a greater risk of developing hypertension and are more likely to have a heart attack or stroke.

3 Pancreas
Sleeping less and at odd hours disrupts the body's metabolism and increases the chances of developing obesity, insulin resistance, and type 2 diabetes.

Immune system
Sleep deprivation increases the odds of catching the common cold and decreases the efficacy of some vaccines.

Life span
People who routinely sleep less than 6 hours each night are 12% more likely to die prematurely than those who sleep 6 to 8 hours, according to one large review study.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Recent studies have firmed up the link between poor sleep and cardiovascular disease in people of color. A 2017 paper published in the *Proceedings of the National Academy of Sciences* showed that among 426 people in three U.S. cities, Black people had a 12.5% higher cardiometabolic risk—the chances of having a cardiovascular event, such as a heart attack or stroke, or developing diabetes—than white people. They also slept 40 fewer minutes per night than white people and spent 10% less time asleep while in bed, a measure of sleep efficiency. The researchers estimated that together, those factors accounted for more than half of the racial difference in cardiometabolic risk.

Poor sleep may also contribute to worse outcomes in Black people with heart problems, a 2018 study found. Black patients who slept fewer than 7 hours at night in the month after a heart attack or episode of unstable angina were at greater risk of another event—or dying—compared with Black people who slept more. Reduced sleep did not have the same detrimental effect for white people with a similar cardiac health history.

The health impacts aren't limited to cardiovascular disease. When we don't sleep enough or our sleep-wake cycle is disrupted, hormone imbalances prompt hunger, which can cause overeating and could contribute to obesity, Johnson says. Researchers have also found that disturbed and insufficient sleep causes alterations in glucose metabolism that can contribute to insulin resistance, a precursor to type 2 diabetes. Obesity and diabetes already disproportionally affect people of color, and researchers suspect sleep disturbances could be making those conditions worse.

Sleep apnea could be having dire health consequences among Hispanic and Black people, says physician-scientist Alberto Ramos at the University of Miami. Large epidemiological studies have suggested apnea and related breathing problems are twice as likely in Hispanic people compared with white people—and often go undiagnosed. In an observational study with 14,440 Hispanic people, Ramos and colleagues found less than 1% of those with apnea had a proper diagnosis.

Poor sleep and sleep apnea have been linked to neurodegenerative diseases—especially Alzheimer's disease—in the general population, Ramos notes. And, he adds, "We know that Hispanics have a two to four times higher risk for dementia." Sleep apnea is one possible contributor to elevated dementia rates and cardiovascular disease—racism and neighborhood noise and air pollution are others—that Jean-Louis and Ramos plan to investigate in the varied Black and Hispanic communities in Miami. At the recent event at Trinity Baptist Church, Jean-Louis recruited 10 new participants.

AFTER JEAN-LOUIS MOVED from San Diego to New York City in 2000 to work at the State University of New York Downstate Medical Center in the heart of Brooklyn, he started to look for ways to engage with the many communities of color near his lab. But he learned the hard way that being Black himself did not mean automatic rapport. When Jean-Louis, a native of Haiti, showed up at a neighborhood barbershop with a clipboard and started to ask people about their sleep habits, they didn't take him seriously. He soon learned to first build trust with the barbershop owners, who would then convince their clients to open up.

Once they did, Jean-Louis found many misconceptions. For instance, he learned many people in these communities do not think of sleep apnea as a problem. "In their minds snoring is a good thing—it means you're in deep sleep," he says. But in fact, snoring is a sign that you're not breathing well during slumber.

Later, when he moved to NYU, he sought other trusted intermediaries. He and Seixas teamed up to recruit volunteers, nominated by their own communities, to be "certified sleep educators." After an 8-week training program on sleep health and counseling, the educators provided tailored advice and emotional support to people in their community, and convinced many to get screened and treated for sleep apnea. During the pandemic, the educators helped convey accurate scientific information about COVID-19—an unintended benefit that demonstrates how trusted they've become, Jean-Louis says. "It was a great surprise," he says. "They were wonderful."

Such educators can deliver tailored and culturally relevant information. "A lot of the guidelines and solutions do not take into account the lived experience of marginalized communities," Alcántara says. The mostly white experts who recommend a certain amount of sleep, Seixas says, "blindly accept and assume that people

can get that sleep." Recommendations have to be applicable to everyone, even a single mom who works two jobs and has two kids, he says.

For a study published last year, Jean-Louis and his team created a website that features Black people sharing their experiences with sleep apnea. "They can see themselves through those stories," Jean-Louis says. "If your brochures only depict white men and women … [Black people] don't relate to this, they don't even open it up." In a randomized trial, the researchers found that participants who used the website reported feeling more able to seek sleep apnea evaluation and adhere to treatment, which involves a machine that delivers pressurized air through a face mask to keep the airway open.

Jean-Louis says Black patients sometimes don't feel welcome in sleep clinics and mistrust medicine and physicians, so he advocates testing these patients at home. Seixas is developing a kit with several wearable devices to screen patients of color for sleep disorders, upload their data, and then provide tailored sleep health advice.

Alcántara, meanwhile, is working to improve access to insomnia therapy for Hispanic people. Therapy for Spanish speakers "is virtually nonexistent," she says. In a review published earlier this year, Alcántara and her colleagues found that of 8182 randomized clinical trials of behavioral interventions to improve sleep, only 7% targeted underserved groups, and only one included Spanish in the treatment.

Cognitive behavioral therapy for insomnia (CBT-I) is the gold-standard treatment for this sleep disorder. It involves a therapist who helps the patient identify thoughts, feelings, and behaviors that interfere with sleep, and suggests ways to mitigate them. At Columbia, Alcántara is currently adapting a digital version of CBT-I approved last year by the U.S. Food and Drug Administration—the first of its kind—for Spanish-speaking Hispanic people. "Adaptation is more than translation," she says. It involves understanding cultural details and family dynamics, such as which family members Hispanics are most likely to trust on medical matters, she says. Next month, she and her team will begin to enroll participants with insomnia in a trial of the culturally adapted CBT-I therapy.

TWO MONTHS AGO, Jean-Louis moved to Miami to launch a Translational Sleep and Circadian Sciences Center at the University of Miami. His new lab aims to probe one factor that could make Black people more vulnerable to poor sleep: There is some evidence that compared with white people, Black people have shorter daily sleep-wake cycles, more readily disrupted by shift work, light pollution, and other factors. "I've been puzzled by this for about 20 years," Jean-Louis says. "We want to get to the bottom of this." He's also building a new network of pastors and barbers in the area for his outreach work.

But engaging with communities of color is only a small part of what's needed, he and others say. Physicians must also play a role, for example. "The training that physicians receive in medical school about sleep is very, very limited," Brown says. And although the number of sleep specialists has been increasing, "there's not enough [of them] treating the people that need to be treated," Johnson adds. "We need more in the right places."

For the first time, the Department of Health and Human Services earlier this year included improving sleep as one of the main disease prevention goals for the next decade. Brown, who chaired the working group that came up with sleep objectives for the project, called Healthy People 2030, is elated that improving sleep is now a national health priority. She is disappointed, however, that tackling sleep disparities wasn't ultimately included, despite all the evidence she and others presented to decision-makers. But she's not deterred. "I am an eternal optimist," she says. "It may take us until Healthy People 2040, but, you know, we're pushing."

Sleep research is starting to influence policy. In 2019, California became the first state to mandate later school times to accommodate children's and teenagers' circadian rhythms. And research on shift work has informed guidelines in the aviation industry and for health care workers, aimed at mitigating the negative health effects of sleep disruption.

These policy wins don't tackle sleep health disparities specifically, but Johnson is quick to point out steps that could. Raising the minimum wage so people don't have to work multiple jobs and do shift work, implementing noise and light pollution regulations, and improving health care access to people of color through telemedicine are among the ideas Johnson and a colleague proposed in an editorial published last year.

For Alcántara, changing policy to improve the sleep of Black, Hispanic, and other people of color in the country is at the core of fighting structural racism. "If we're thinking about addressing sleep as a social justice issue, then yes, you have to focus on the policies that are driving who has the luxury of sleeping well and who doesn't," she says. ■





Sleep researcher Girardin Jean-Louis recently visited Trinity Baptist Church in North Miami, Florida, to speak with churchgoers about the importance of sleep for overall health and wellness (top). At New York University, Azizi Seixas (bottom, center) uses digital tools to screen people in the Black community for potential sleep health issues.

PHOTOS: (TOP TO BOTTOM) GIRARDIN JEAN-LOUIS; MIKE YAMIN AND AZIZI SEIXAS

REVIEW

The inescapable drive to sleep: Overlapping mechanisms of sleep and sedation

Nicholas P. Franks and William Wisden

Common human experience is that a long period without sleep is unsustainable, and it is also detrimental to health and behavior. The powerful and primal urge to sleep after sleep deprivation is intense and seems inescapable. The longer we stay awake, the more we feel the need to sleep, and however much we resist, we will inevitably succumb. Although it is obvious what benefits derive from other common and strong physiological drives, such as hunger, sex, and thirst, it is less obvious what drives us to sleep and what benefits accrue. Understanding the biochemical or circuit basis for the sleep drive could enable the benefits of sleep to be artificially stimulated with a new generation of sedative drugs.

Why do we sleep? The evolutionary advantages of remaining permanently awake and alert seem so obvious that sleep must provide some essential need. Dolphins, fur seals, and whales alternate their sleep in their cortical hemispheres so that they can keep swimming: One half of the brain "sleeps" while the other is awake (*1*). This observation alone seems to highlight the importance of sleep, at least for mammals. Despite intense research, why we sleep remains one of the most baffling questions in neuroscience. In this article, we discuss why we might sleep, what mechanisms force us to sleep when we are sleep deprived, and whether certain anesthetics and sedatives hijack this natural drive to exert their effects on human consciousness.

Do we have to sleep and if so, why?

Although it is often stated that "all animals sleep," it is not obvious, or proven, that all types of animals sleep or sleep in the same way or for the same purpose. Nevertheless, it seems clear that sleep is good for us humans and is in some way restorative. Moreover, we know from personal experience that missing sleep is detrimental. In the short term, total sleep deprivation decreases cognitive ability and mood (*2*). Twenty-four hours of enforced wakefulness has a greater detrimental effect on driving ability than consuming the legal limit for alcohol (*3*). Therefore, something is clearly deteriorating in our brain function when sleep is missing. What about the long-term effects of sleep loss? A recent study tracked nearly 8000 UK government employees for 25 years, from middle age through to their early seventies. The results showed a higher dementia incidence in people between 50 and 60 years of age who had been chronically short sleepers (6 hours a night or less, self-reported) (*4*). Thus, in the long term, less sleep correlates with increased pathology.

Department of Life Sciences and UK Dementia Research Institute, Imperial College London, London SW7 2AZ, UK.
Correspondence: n.franks@imperial.ac.uk (N.P.F.); w.wisden@imperial.ac.uk (W.W.)

Because we all experience sleep, there is a common understanding of what the sleeping state entails for us subjectively, which we are comfortable extrapolating to other animals, at least to other mammals. Sleep can be measured using the external electrical signatures of brain and muscle (*5*) and can, in vertebrate animals, be categorized as non–rapid eye movement (NREM) and REM sleep, or at least NREM-like and REM-like. However, depending on species, these sleep states differ in their physiological properties and timing (*5*). The detailed properties of any one sleep state, electroencephalographic (EEG) pattern, or calcium-oscillatory activity are unlikely to be critical. Some other common, as yet unknown, property of the sleep state is probably the key ingredient for why sleep is essential. Much as birds and mammals use different circuitry and brain organizations to carry out cognitive tasks with high effectiveness, sleep is likely to be an essential emergent property regardless of how the brain is organized (*6*). The requirement for sleep could originate at the cellular level, with neurons fatiguing with time spent awake, analogous to the metabolic fatigue of skeletal muscle with exercise. But it is not obvious whether all parts of the brain need to sleep; circuits that drive breathing, for example, do not seem to rest.

Notwithstanding the many variations of sleeping, is there a core reason to sleep that is so fundamental that sleep is obligatory? Although there are many activities that occur during sleep but occur less often during waking, few of these activities are likely to be the fundamental and primal reason that we must sleep. A telling fact is that of the hundreds of mouse mutants generated, no mutant mouse has yet been discovered that does not exhibit NREM sleep (*7*). Many mutations affect the depth of NREM sleep or its timing (e.g., circadian mutants), but so far all have NREM sleep (*7*). The same picture emerges with circuit manipulations. Lesions of sleep-promoting centers can permanently reduce the amount of sleep in rodents, yet some sleep always remains (*8*–*10*). So far, no cell or circuit lesion has removed all sleep.

So, what is sleep for? A widespread, but in our view mistaken, belief is that if a detriment in behavior or well-being can be identified as a consequence of sleep deprivation, then this provides evidence for the reason that we sleep. However, it seems more likely that sleep provides some fundamental "housekeeping" or metabolic functions that are necessary for a healthy brain and body, so that there will be multiple deficits of one sort or another that arise when sleep is denied (Fig. 1A). By "housekeeping," we mean basic structural or metabolic processes that are essential for the normal functioning of the brain. The most important constraint when considering what these functions might be is that they cannot, apparently, occur in the waking brain; restful waking does not satisfy our need for sleep. Much like the cleaning teams who move into empty offices during the night and whose work would be almost impossible during the daytime bustle, some essential and restorative process is underway after we slip into sleep, when normal brain function is at least partly suspended. The need for unconsciousness is perhaps the key thing to explain.

Mechanisms that stand out for being incompatible with consciousness would be those that involve structural changes in the brain. For example, the glymphatic hypothesis, that NREM sleep allows the clearance of metabolites from the brain (*11*), posits that the end feet of astrocytes shrink to enhance glymphatic flow (*11*). Similarly, any remodeling of synapses to rebalance network activity acquired during wakefulness might preclude cognitive activity or awareness and might only be possible "offline" (*12*). Another factor to consider is temperature. During sustained NREM sleep, brain neocortex temperature in mice drops by ~2°C (*13*, *14*) (Fig. 1B). In fact, specific hypothalamic circuitry exists to cool the brain and simultaneously induce NREM sleep (*15*). It has been proposed that a key function of sleep is to preserve energy (*16*). For humans, however, this temperature decrease with sleep would result in only a modest saving in metabolic energy, equivalent to about two slices of bread (*16*). A lower temperature with sleep could nonetheless be required for an unknown restorative process. This was first suggested >30 years ago (*17*) and might be linked to synaptic remodeling. Cooling during NREM sleep induces the expression of the cold-inducible RNA-binding protein and RNA-binding motif protein 3 genes (*13*), and the proteins encoded by these genes could be required for structural remodeling (they are also induced in some hibernators). Furthermore, cooling even by only a degree or two would prolong inhibitory postsynaptic currents by about the same extent as do many general anesthetics at sedative doses (*18*) (Fig. 1B); this would make normal cognition essentially impossible, so a sleeping state would be necessary for these processes to occur. Finally,

the delta and theta electrical oscillations that occur during NREM and REM sleep could be important for rebalancing the network during sleep (*19*), and such oscillations may also be incompatible with normal cognitive function. A good example of how oscillations can block cognition is provided by the anesthetic ketamine. This compound elicits a highly selective, delta-like rhythm (1 to 3 Hz) in layer 5 pyramidal neurons of the retrosplenial cortex, thus inducing a dissociation of stimulus detection and response (*20*), which is obviously inconsistent with a normally functioning waking state.

In contrast to NREM sleep, REM sleep may not be essential for life. It is possible to genetically ablate all REM sleep in laboratory mice with no apparent ill effects (*21*). On the other hand, and perhaps unexpectedly, people who report good and refreshing sleep have good REM sleep continuity, whereas consolidated NREM appears to be less important for the perception of a good night's sleep (*22*). Theories for the function of REM sleep are not constrained by the need for the brain to be offline because the brain is active during REM (*5*) and brain temperature rises (*13*). It could be that because REM sleep always follows NREM sleep, their functions are linked. For example, REM could be an algorithm that tests whether the restorative function of NREM has been completed. If it has, we wake up. An important question to solve is, what determines the timing and sequence of the NREM and REM cycles?

The homeostatic drive to sleep

In many animal species, sleep deprivation increases the drive for both NREM and REM sleep (*2*). This homeostatic sleep drive strengthens the idea that some important restorative process is underway and is the reason that we must "catch up" on lost sleep. This catching up, however, is not just about quantity but also quality, at least for NREM sleep. If we miss one night's sleep of 7 hours, say, our next night of sleep is likely to be subjectively "deeper," as well as potentially longer. Although it remains mysterious what this deep sleep is providing physiologically, a pervasive concept is that a reliable marker of NREM sleep depth in both humans and other terrestrial vertebrates is the power exhibited in the EEG in the delta range of frequencies (~0.5 to 4 Hz). The EEG delta power, particularly the higher end of this band of frequencies (*14*), markedly increases after sleep deprivation but rapidly rebounds. However, many hours more are needed for the actual sleep loss to be restored, implying that additional, potentially restorative processes that are not reflected in EEG delta power per se are underway (Fig. 2). The high delta power immediately after sleep deprivation might nonetheless be permissive for the restorative benefits that follow, perhaps through brain cooling.

We can consciously fight sleepiness to a certain extent, a top-down control from the neocortex imposing its will on the "sleep centers," and we might be determined to stay awake, but eventually we give way to sleep. After a certain period of intense cortical activity, local delta oscillations "break out" in areas of the neocortex even if animals are behaving as if they are awake (*23*). One might then ask, if this restricted "sleep" is providing restorative benefits locally, then why does the whole animal need to shut down and enter global sleep? Why could the brain not be continually having smaller domains entering "sleep" so that normal function is still possible at the whole-animal level? This could be because it is necessary to cool the brain for the restoration to occur, and this cannot be achieved locally, only centrally. In our view, the term "sleep" is best restricted to a whole-animal behavior, and local short periods of increases in delta power are clearly insufficient for whatever purpose global sleep serves.

The sleep drive can also be assessed independently of EEG delta power by changes in behavior. When we are strongly sleep deprived, we become highly motivated to find a way to sleep, in the same way that strong thirst and hunger motivate us to drink and eat. If sleep deprived enough, we will do almost anything to sleep. Afternoon naps provide additional evidence for an accumulating sleep drive during the time spent awake and show that this drive can be partially dissipated by short periods of acute sleep. An afternoon nap delays the onset of later NREM sleep, which starts off with a lower EEG delta power than it would have done had no nap been taken (*2*). If naps are too long, it can be difficult to get to sleep at all in the evening. This implies that there must be some feedback between the process(es) that provides restorative benefits and the mechanism that tracks the time awake.

Classical views of how the hunger drive is organized implicated hunger centers in the hypothalamus; for the sleep drive, the preoptic hypothalamus was shown to have a particularly strong sleep-promoting role (*24*). However, it is now clear that motivational drives for food and water emerge in a distributed way throughout the brain (*25*). Similarly, the basis for the sleep drive is likely to be more distributed than originally anticipated (Fig. 3A). Indeed, in mice, the circuitry that induces both NREM and REM sleep is itself widely distributed (*26*). It could

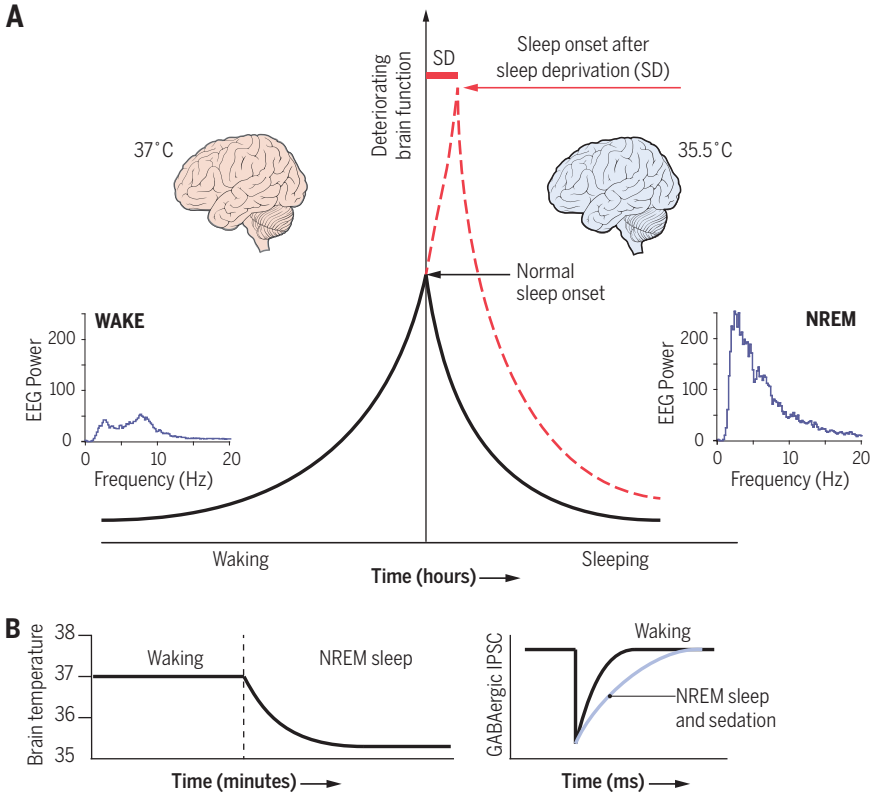


Fig. 1. The most probable reason for sleep is to allow fundamental housekeeping processes to occur. (**A**) During waking, there is a progressive deterioration of brain function until a point is reached when sleep is triggered, consciousness is lost, and repair is initiated. This change in vigilance state is reflected in the EEG power spectrum, which moves from a low-power broad spectrum to a high-power spectrum that peaks at delta frequencies (0.5 to 4 Hz) and is characteristic of NREM sleep. With prolonged NREM sleep, the brain cools. If sleep is delayed, brain deterioration continues until a point is reached when sleep onset is inevitable. (**B**) During sustained NREM sleep, the brain cools by ~1.5°C, which would prolong inhibitory postsynaptic currents by about the same extent as sedative doses of anesthetics.

well be that all parts of the sleep circuitry distributed throughout the brain can implement homeostasis. Hypothalamic preoptic galanin neurons (*27*, *28*), basal forebrain cholinergic neurons (*29*), midbrain ventral tegmental area (VTA) GABA neurons (*30*), raphe serotonin neurons (*31*), neocortical layer 5 pyramidal neurons (*10*), and astrocytes (*32*) have so far been proposed as mediators of sleep homeostasis. It is unclear what "tiredness signals" are being sensed by this distributed circuitry (see below).

Because NREM sleep can be dispensed with temporarily, for example, if we stay up all night, this implies that any restorative function of NREM sleep need not be immediate (e.g., mice and humans catch up on lost sleep over many hours, if not days). Similarly, frigate birds typically sleep very little (0.7 hours/day) during 10-day periods on the wing yet sleep nearly 13 hours per day when on land (*33*). Thus, whatever is being tracked during sleep deprivation and leading to the drive to sleep can be partially overridden (*34*); therefore, this is not the same as faster breathing and heart rate after exertion to immediately resupply oxygen. Hunger and thirst are also drives that can be similarly put on hold by the executive control centers in the brain, at least for a while. A gradual process of recovering lost sleep over many hours suggests anabolic or catabolic housekeeping processes, i.e., synthesis or degradation.

A better understanding of the sleep drive and how the need for sleep is "clocked" might reveal the functions of sleep. One way that sleep drive might track wakefulness has been discovered in fruit flies and involves a changing redox state with time spent awake and a change in the activity of potassium channels, resulting in an increased firing rate of sleep-promoting neurons (*35*). In principle, such a mechanism could work in mammals too. In vertebrate species, however, the actual biochemical processes that track the time spent awake and lead to increased NREM sleep pressure remain unclear. The simple and compelling idea that there is a substance that accumulates during sleep, tracks time awake, provides feedback to induce sleep, and is then degraded during sleep has been proposed many times, starting with Rosenbaum in 1892 (*36*). Adenosine is the best-known candidate molecule: It accumulates with time spent awake in a few select brain areas, such as the basal forebrain (*37*), where it (or adenosine trisphosphate) could be secreted from cholinergic neurons in the basal forebrain and astrocytes (*29*, *38*). In addition to its many metabolic roles, adenosine can act through metabotropic receptors to induce NREM sleep (*36*), but appealing though this idea is, definitive proof that adenosine is the primary cause of the sleep drive is lacking, especially because the adenosine increases that correlate with prolonged waking seem to happen in only a few brain regions, yet the sleep homeostasis circuitry is widely distributed. Many other possible sleep-inducing factors have been identified, including interleukin 1β and tumor necrosis factor–α, the levels of which increase in the neocortex during sleep deprivation (*6*). The brain is not the whole story. Certainly, peripheral tissue can signal sleep; unidentified systemic factors from skeletal muscle contribute to the sleep drive (*39*), perhaps explaining why sleepiness can often be induced by exercise (*2*).

Phosphorylation tracks sleep need. In mice, a gain-of-function mutation in the salt-inducible kinase 3 (*Sik3*) gene reduces NREM sleep homeostasis as defined by delta power and increases the amount of sleep (*7*). This particular kinase is widely expressed, including in peripheral organs, and could also act throughout the brain, in keeping with the distributed sleep homeostasis–promoting circuitry. It is not known how wakefulness activates Sik3, but the targets of this kinase are numerous. In mice, as wakefulness progresses, there are steady increases in phosphorylation (and dephosphorylation) of many proteins in the forebrain, including ion channels (*40*, *41*), which might increase or decrease the firing rate of sleep-promoting neurons or inhibit wake-promoting neurons.

One generally ignored question is, what determines when we wake up? Although this is partly a circadian influence, waking from a refreshing night's sleep suggests that the homeostatic process of sleep has been completed. Therefore, presumably, the restorative process of sleep is continuously monitored to determine that the brain is now ready to wake. This could be related to the NREM-REM cycle discussed above because we tend to wake from REM sleep.

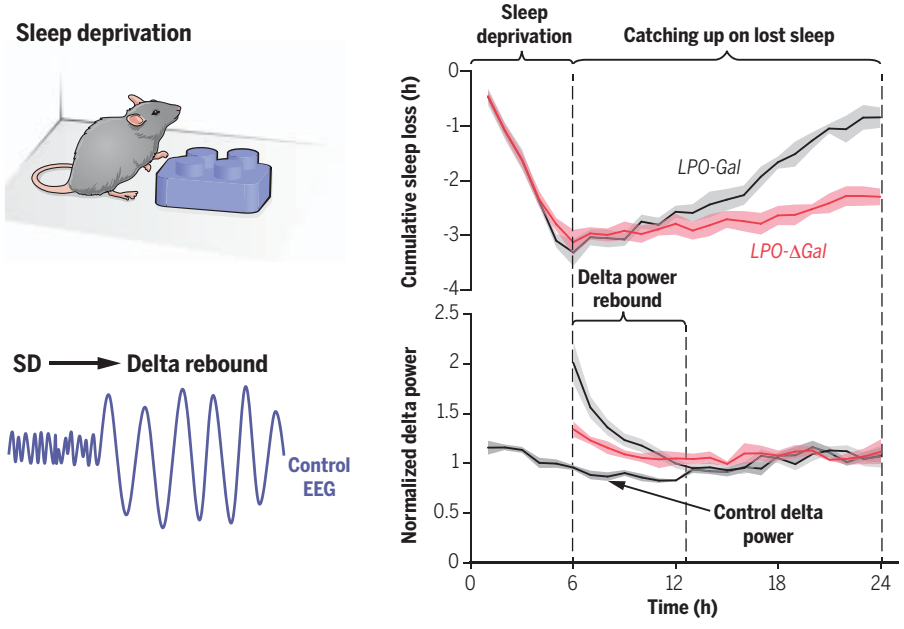


Fig. 2. After sleep deprivation with novel objects, delta oscillations in mice rebound rapidly, but recovering lost sleep takes much longer. After 6 hours of sleep deprivation using novel objects, lost sleep is recovered over the following 24 hours, but the rebound in delta oscillations comes back to baseline after only about 6 hours. Selective genetic ablation of certain neuronal populations can greatly decrease sleep homeostatic responses. For example, if galanin neurons are selectively ablated in the lateral preoptic hypothalamus, both measures of sleep homeostasis are significantly blunted. Graphs are replotted from data in (*27*) with permission under the Creative Commons CC-BY license.

Can drugs provide the restorative benefits of natural sleep?

The drive to sleep after a certain amount of sleep deprivation is so intense that it resembles drug-induced sedation. "I'm just going to give you something to make you sleep" is an almost universally used and comforting metaphor when we undergo an investigative procedure or are being prepared for surgery. This is probably more than a metaphor. Although deep anesthesia suitable for surgery might be best described as a reversible coma, lower concentrations of anesthetics produce a sedative state. The overall pattern of brain activity in humans, as measured using functional magnetic resonance imaging, is markedly similar during NREM sleep and after sedation with dexmedetomidine or propofol (Fig. 3B) (*42*). Humans given the α2 adrenergic agonist dexmedetomidine enter a state resembling stage 3 NREM sleep, the deepest type of sleep (*43*). We now know that this is because sedatives and anesthetic agents can act in the sleep-wake circuitry (*18*, *44*–*46*), especially in the midbrain circuitry comprising the lateral habenula, VTA, and hypothalamus, to mimic NREM sleep (although there are no drugs known that mimic REM sleep). For example, the preoptic hypothalamic galanin neurons that contribute to NREM sleep homeostasis (*27*, *28*) partly mediate

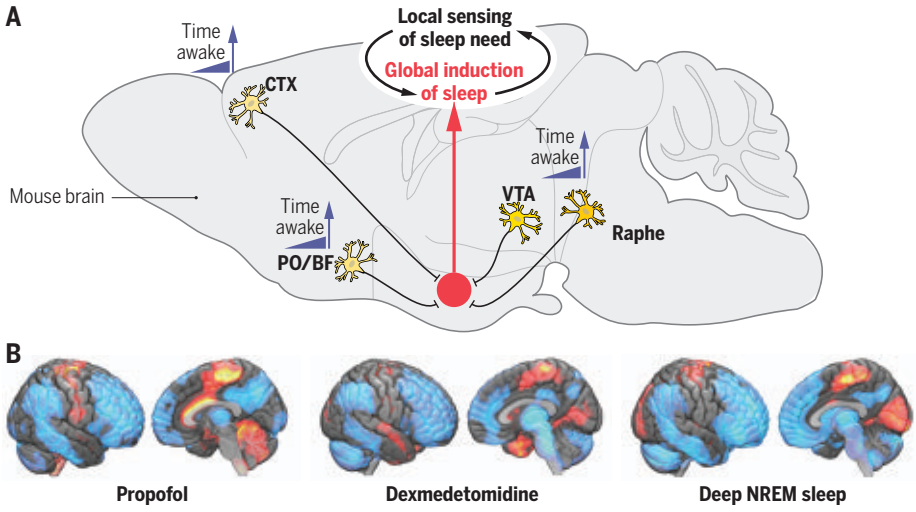


Fig. 3. Distributed circuitry senses local sleep need, which triggers global sleep, and sedatives hijack some of these pathways. (**A**) Current data support a model in which there is distributed circuitry that locally senses the need to sleep and then sleep is initiated globally, perhaps by the lateral hypothalamus (red sphere). In addition to this circuitry that registers "tiredness," signals from glia and muscle also contribute to the sleep drive. (**B**) Overall brain activity in humans during deep NREM sleep is very similar to that caused by general anesthetics and sedatives. Images show changes in regional cerebral blood flow measured with positron emission tomography. Figure is adapted from (*42*) with permission under the Creative Commons CC-BY license. BF, basal forebrain; CTX, neocortex; PO, preoptic hypothalamic area.

the arousable NREM-like sleep induced by dexmedetomidine (*27*). Propofol and isoflurane both induce a sedative state (*47*, *48*), in part by activating the lateral habenula, a midbrain nucleus important for ensuring NREM sleep stability (*47*). Selectively activating GABA neurons by chemogenetics in the VTA also produces a particularly deep, sedative-like NREM state for 4 to 5 hours (*9*). The activation of these VTA GABAergic neurons enhances isoflurane and sevoflurane anesthesia (*49*, *50*), and their inhibition promotes arousal from anesthesia (*49*). These VTA GABA neurons are also needed for sleep homeostasis (*30*). Therefore, these particular cells, or a subset of them, could be a potential target for new sedatives that provide restorative benefits.

What is the evidence that current sedatives provide any restorative benefits? As many of us have experienced, recovery from surgery or a same-day procedure can be exhausting and feels the opposite of restorative. However, the trauma experienced during these operations could easily overwhelm any beneficial restorative benefits that the sedatives themselves might provide. More importantly, because we do not know what these restorative benefits of sleep are, we have no way of directly measuring whether they accrue during sedation. Instead, we have to rely on other measures such as the EEG delta power; however, as mentioned above, this is more likely to correlate with sleep depth rather than any restorative processes. An additional confound is that many drugs will affect the shape and timing of oscillations in the EEG independently of what drives them. For example, the anesthetics propofol and isoflurane potentiate the actions of GABA at $GABA_A$ receptors (*18*), and this will reduce EEG frequency independently of sleep. Therefore, a direct comparison between the EEG during sleep and during sedation is not straightforward with these drugs. A better approach is to determine whether sedative drugs can "repay" sleep debt after sleep deprivation. Some anesthetics can repay NREM debt but few repay REM debt (*51*). Even if current sedatives act on natural sleep pathways, they certainly have other targets. The receptors are not restricted to sleep-inducing pathways. Moreover, the natural pattern of NREM-REM cycles does not occur during sedation, so current drugs clearly do not exactly mimic natural sleep. Therefore, until a drug is developed that selectively acts on sleep pathways, we cannot be sure that the restorative benefits of sleep can be provided artificially by drugs.

Conclusions

The overwhelming urge to sleep shapes the lives of all humans. The most likely reason we sleep is to allow fundamental housekeeping processes to occur, and these processes require an unconscious brain. The lower brain temperatures during sustained NREM sleep would slow inhibitory postsynaptic currents by about the same extent as sedatives and thus be inconsistent with consciousness. Multiple deficits in brain function result if sleep is denied; the time course of catching up on lost sleep implies anabolic or catabolic processes. Sleep need is sensed locally by circuitry distributed throughout the brain, and this circuitry eventually triggers global sleep. In principle, drugs could be developed that more selectively target the restorative properties of natural sleep.

REFERENCES AND NOTES

1. O. I. Lyamin, L. M. Mukhametov, J. M. Siegel, *Curr. Opin. Neurobiol.* **44**, 144–151 (2017).
2. D. J. Dijk, H. P. Landolt, *Handb. Exp. Pharmacol.* **253**, 441–481 (2019).
3. J. Lowrie, H. Brownlow, *BMC Public Health* **20**, 980 (2020).
4. S. Sabia *et al.*, *Nat. Commun.* **12**, 2289 (2021).
5. M. S. Blumberg, J. A. Lesku, P.-A. Libourel, M. H. Schmidt, N. C. Rattenborg, *Curr. Biol.* **30**, R38–R49 (2020).
6. J. M. Krueger, J. T. Nguyen, C. J. Dykstra-Aiello, P. Taishi, *Sleep Med. Rev.* **43**, 14–21 (2019).
7. H. Funato *et al.*, *Nature* **539**, 378–383 (2016).
8. J. Lu, M. A. Greco, P. Shiromani, C. B. Saper, *J. Neurosci.* **20**, 3830–3842 (2000).
9. X. Yu *et al.*, *Nat. Neurosci.* **22**, 106–119 (2019).
10. L. B. Krone *et al.*, *Nat. Neurosci.* **24**, 1210–1215 (2021).
11. M. Nedergaard, S. A. Goldman, *Science* **370**, 50–56 (2020).
12. C. Cirelli, G. Tononi, *Semin. Cell Dev. Biol.* 10.1016/j.semcdb.2021.02.007 (2021).
13. M. M. Hoekstra, Y. Emmenegger, J. Hubbard, P. Franken, *eLife* **8**, e43400 (2019).
14. J. Hubbard *et al.*, *Nat. Commun.* **11**, 3130 (2020).
15. E. C. Harding *et al.*, *Curr. Biol.* **28**, 2263–2273.e4 (2018).
16. C. M. Jung *et al.*, *J. Physiol.* **589**, 235–244 (2011).
17. D. McGinty, R. Szymusiak, *Trends Neurosci.* **13**, 480–487 (1990).
18. N. P. Franks, *Nat. Rev. Neurosci.* **9**, 370–386 (2008).
19. A. R. Adamantidis, C. Gutierrez Herrera, T. C. Gent, *Nat. Rev. Neurosci.* **20**, 746–762 (2019).
20. S. Vesuna *et al.*, *Nature* **586**, 87–94 (2020).
21. Y. Niwa *et al.*, *Cell Rep.* **24**, 2231–2247.e7 (2018).
22. C. della Monica, S. Johnsen, G. Atzori, J. A. Groeger, D.-J. Dijk, *Front. Psychiatry* **9**, 255 (2018).
23. V. V. Vyazovskiy *et al.*, *Nature* **472**, 443–447 (2011).
24. C. B. Saper, T. E. Scammell, J. Lu, *Nature* **437**, 1257–1263 (2005).
25. J. E. Allen, T. E. Sutherland, *Science* **364**, 738–739 (2019).
26. D. Liu, Y. Dan, *Annu. Rev. Neurosci.* **42**, 27–46 (2019).
27. Y. Ma *et al.*, *Curr. Biol.* **29**, 3315–3322.e3 (2019).
28. S. Reichert, O. Pavón Arocas, J. Rihel, *Neuron* **104**, 370–384.e5 (2019).
29. A. V. Kalinchuk, T. Porkka-Heiskanen, R. W. McCarley, R. Basheer, *Eur. J. Neurosci.* **41**, 182–195 (2015).
30. X. Yu *et al.*, *Mol. Psychiatry* 10.1038/s41380-020-0810-9 (2020).
31. G. Oikonomou *et al.*, *Neuron* **103**, 686–701.e8 (2019).
32. A. M. Ingiosi *et al.*, *Curr. Biol.* **30**, 4373–4383.e7 (2020).
33. N. C. Rattenborg, *Interface Focus* **7**, 20160082 (2017).
34. A. Eban-Rothschild, W. J. Giardino, L. de Lecea, *Curr. Opin. Neurobiol.* **44**, 132–138 (2017).
35. A. Kempf, S. M. Song, C. B. Talbot, G. Miesenböck, *Nature* **568**, 230–234 (2019).
36. M. Lazarus, J.-F. Chen, Z.-L. Huang, Y. Urade, B. B. Fredholm, *Handb. Exp. Pharmacol.* **253**, 359–381 (2019).
37. T. Porkka-Heiskanen *et al.*, *Science* **276**, 1265–1268 (1997).
38. M. G. Frank, *Handb. Exp. Pharmacol.* **253**, 83–96 (2019).
39. J. C. Ehlen *et al.*, *eLife* **6**, e26557 (2017).
40. F. Brüning *et al.*, *Science* **366**, eaav3617 (2019).
41. Z. Wang *et al.*, *Nature* **558**, 435–439 (2018).
42. A. Scheinin *et al.*, *J. Neurosci.* **41**, 1769–1778 (2021).
43. O. Akeju *et al.*, *Clin. Neurophysiol.* **129**, 69–78 (2018).
44. H. C. Hemmings Jr *et al.*, *Trends Pharmacol. Sci.* **40**, 464–481 (2019).
45. L. F. Jiang-Xie *et al.*, *Neuron* **102**, 1053–1065.e4 (2019).
46. M. B. Kelz, G. A. Mashour, *Curr. Biol.* **29**, R1199–R1210 (2019).
47. C. Gelegen *et al.*, *Curr. Biol.* **28**, 580–587.e5 (2018).
48. C. Liu *et al.*, *Front. Mol. Neurosci.* **14**, 628996 (2021).
49. L. Yin *et al.*, *Front. Neural Circuits* **13**, 73 (2019).
50. K. Vlasov *et al.*, *Anesth. Analg.* **132**, e50–e55 (2021).
51. D. Pal, W. J. Lipinski, A. J. Walker, A. M. Turner, G. A. Mashour, *Anesthesiology* **114**, 302–310 (2011).

ACKNOWLEDGMENTS

We thank D.-J. Dijk (University of Surrey, UK) for helpful discussion. **Funding:** Work in the Franks-Wisden laboratory is funded by the Wellcome Trust (grants 107839/Z/15/Z and 107841/Z/15/Z) and the UK Dementia Research Institute (grant UK DRI-5004). **Competing interests:** The authors declare no competing financial interests.

10.1126/science.abi8372

REVIEW

Brain neural patterns and the memory function of sleep

Gabrielle Girardeau[1]* and Vítor Lopes-dos-Santos[2]

Sleep is crucial for healthy cognition, including memory. The two main phases of sleep, REM (rapid eye movement) and non-REM sleep, are associated with characteristic electrophysiological patterns that are recorded using surface and intracranial electrodes. These patterns include sharp-wave ripples, cortical slow oscillations, delta waves, and spindles during non-REM sleep and theta oscillations during REM sleep. They reflect the precisely timed activity of underlying neural circuits. Here, we review how these electrical signatures have been guiding our understanding of the circuits and processes sustaining memory consolidation during sleep, focusing on hippocampal theta oscillations and sharp-wave ripples and how they coordinate with cortical patterns. Finally, we highlight how these brain patterns could also sustain sleep-dependent homeostatic processes and evoke several potential future directions for research on the memory function of sleep.

Memory formation is the challenging process of selecting which new experiences will be stored and integrated into an existing structure of memories that needs to be simultaneously preserved and modified. During wakefulness, this occurs concurrently with an uninterrupted flow of new sensory experiences. Sleep provides a window of opportunity for the brain to sort and reinforce newly encoded memories in absence of the incessant barrage of external information. This process, called consolidation, leads to the generation of long-lasting memory traces or engrams whose activation during wakefulness supports the recall of information.

During sleep, a myriad of neural networks involved in memory processing are endogenously activated. Their activity generates electrical potentials captured using noninvasive surface electrodes [electroencephalograms (EEGs)] or intracranial electrodes that can record local field potentials (LFPs) as well as action potentials (spiking activity). A large amount of effort has been devoted to describing how we can use meaningful patterns in these electrical fluctuations to understand the brain. These patterns include oscillations (e.g., theta rhythm), transient potentials with an identifiable waveform (e.g., dentate spikes), and spiking activity patterns (e.g., up and down states). Combining signal analysis and anatomical data, as well as targeted intracranial recordings or manipulation of superficial and deep structures, has boosted our understanding of the cellular basis of these patterns. Ultimately, these advances may lead to an understanding of the role that sleep brain patterns play in learning and memory.

[1]Institut du Fer a Moulin, UMR-S 1270 INSERM and Sorbonne Université, 75005 Paris, France. [2]Medical Research Council Brain Network Dynamics Unit, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford OX1 3TH, UK.
*Corresponding author. Email: gabrielle.girardeau@inserm.fr

Non–rapid eye movement sleep and hippocampal sharp-wave ripples

One of the most important patterns in sleep is the sharp-wave ripple (SWR) complex (Fig. 1). The hippocampus is a three-layer structure in which the information flows from the dentate gyrus to the CA1 region through CA3. During sleep, CA3 pyramidal neurons spontaneously activate in synchronous bursts that trigger a massive activation of CA1 pyramidal cells. In the stratum radiatum, the CA3 input on pyramidal cell dendrites creates the sharp wave, whereas in the CA1 pyramidal cell layer, the interplay between activated pyramidal cells and interneurons gives rise to the fast (100 to 250 Hz) oscillatory part of the event: the ripple (*1*). The two-step theory (*2*) postulates that first, a subgroup of CA3 and CA1 cells are coordinated by theta oscillations during an experience and form cell assemblies encoding the corresponding new information. Then, in subsequent sleep periods, these CA3 assemblies spontaneously ignite SWR events that reactivate the associated CA1 ensembles and promote the strengthening of their connections, which ultimately leads to memory consolidation. Consistent with this theory, pairs of CA1 pyramidal cells that cofire during the exploration of an open field maintain this correlation during subsequent sleep SWRs (*3*). The persistence of the activity correlations observed in awakening in subsequent sleep is commonly referred to as sleep reactivation. By using a wide range of methods (*4*, *5*), subsequent studies established that cofiring patterns and entire sequences of place cells that are activated during wakefulness are reinstated during the SWRs of the following sleep epoch ["replay" (*6*); Fig. 1]. Importantly, reactivation was also shown in humans (*7*).

The first causal studies for the role of reactivation in memory consolidation developed closed-loop paradigms (Fig. 2) to disturb sleep ripples and therefore the associated reactivation. They showed drastic spatial memory impairment (*8*, *9*). Optogenetic silencing of CA1 pyramidal neurons during sleep SWRs after

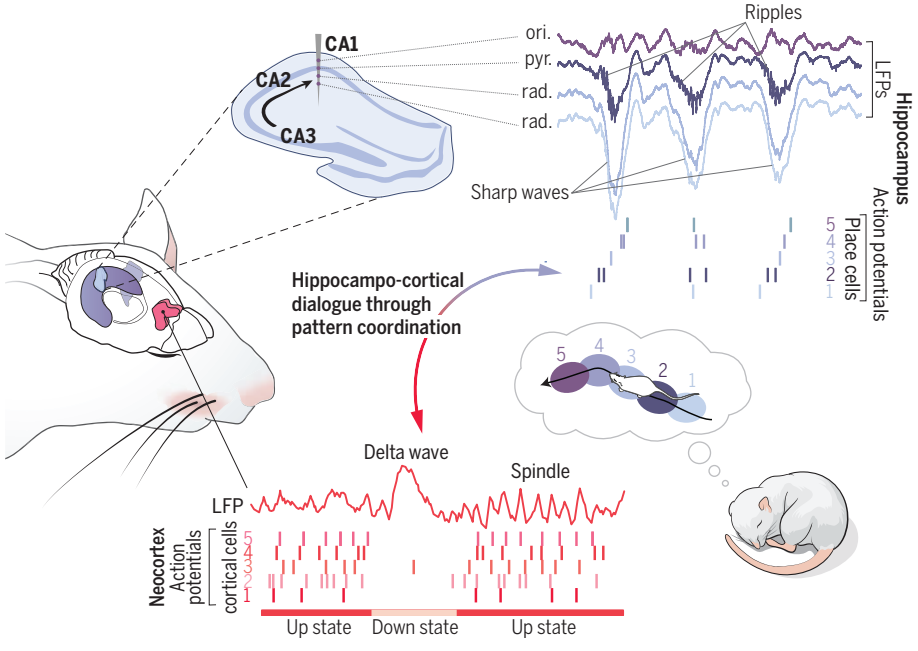


Fig. 1. Hippocampal and cortical patterns coordinate during NREM sleep to sustain memory consolidation. In the hippocampus, coordinated input from CA3 depolarizes CA1 pyramidal neurons to create a sharp wave in the radiatum layer (rad.) that reverses in the oriens layer (ori.) and a fast, 200-Hz ripple in the pyramidal layer (pyr.). SWRs are associated with place cell activity that recapitulates the trajectories experienced in the previous wakefulness epoch. In the neocortex, unit activity alternates between periods of high activity (up state) associated with spindles and silence (down state) reflected on the LFP as a delta wave.

the exploration of novel environments impairs the reinstatement of these cell assemblies upon reexposure to the same environment, suggesting that memory impairments at recall are due to a lack of consolidation of the spatial maps, or engrams, sustaining the memory (*10*). Various factors influence SWR-associated reactivation during sleep. For example, reactivation in CA1 is stronger, and lasts longer (*11*), after novelty. It is also biased toward the activity previously expressed in theta cycles associated with strong mid-gamma oscillations (50 to 100 Hz), suggesting that assemblies formed during the heightened influence of the entorhinal cortex, which is thought to convey new extra-hippocampal information, are preferentially reactivated (*12*). Most SWR and reactivation studies focus on CA1. However, social memory traces are reactivated in CA2 during SWRs, and their bidirectional modulation enhances or impairs social memory (*13*). These results suggest that although CA3 might bias the SWR reactivated assemblies to consolidate spatial memories, CA2 is essential to bias SWR content toward social memories.

The development of algorithms for fast, online detection of specific replay content, as opposed to the mere detection of ripples on LFPs, is a necessary step to further our understanding of the role of sleep replay. Along this line, Gridchyn *et al.* (*14*) trained rats to forage in two environments and disrupted the following sleep and rest SWR events, except the ones reactivating the first environment. The performance on this environment was better than on the second one, indicating that the consolidation of the spatial memories related to the first environment were spared from disruption. Altogether, the results accumulated over the past decades strongly indicate that reactivation of hippocampal ensembles associated with novel information and learning during sleep SWRs is essential for memory consolidation. Surprisingly, however, it is still unknown whether hippocampal reactivation also occurs in the ventral part of the hippocampus, which has different connectivity and is involved in stress and anxiety. In addition, hippocampal dentate spikes that reflect strong cortical inputs to the dentate gyrus during non–rapid eye movement (NREM) sleep have been identified as potential players in the NREM consolidation processes but remain to be further explored (*15*).

Although this review focuses on sleep, SWRs also occur during awake immobility and nonexploratory behaviors (grooming, eating, etc.). There are no clear qualitative differences between awake and sleep ripples, but their replay content differs (*6*). A major challenge will be to understand whether and how NREM sleep background (neuromodulation, reduced external inputs, cortical and subcortical NREM-specific activity, etc.) makes sleep ripples and their associated neuronal content functionally different from the awake ones. Further, these differences could either be characterized as a simple sleep-wake dichotomy or occupy a multidimensional functional space (consolidation, forgetting, planning, memory reorganization, decision-making, etc.) depending on numerous parameters, including neuromodulatory levels, attention or alertness, ongoing behavior, sleep debt, circadian rhythm, consolidation needs, immediate and long-term previous experience, or NREM sleep substages.

Hippocampo-cortical coordination through NREM sleep patterns

All major theories for long-term memory consolidation involve communication between the hippocampus and the neocortex (*16*). During NREM sleep, cortical circuits undergo an alternation of periods of marked high and low population activity, referred to as up and down states, respectively. This alternation translates in LFPs as the NREM sleep canonical slow oscillation. In particular, down states are associated with distinctive LFP deflections called delta waves. Delta waves are often followed by spindles, which are bouts of 10- to 15-Hz oscillations originating from the thalamus. All of these cortical rhythms have, individually, but mostly through their coordination with other hippocampal and cortical patterns, been related to memory consolidation (*16*–*18*) (Fig. 2).

Transcranial stimulation in humans can be used to boost slow oscillations during NREM sleep, and the manipulation enhances performance at retrieval on the next day (*19*). Numerous EEG correlational studies have highlighted the importance of slow waves and spindles for memory consolidation (*16*). In rodents, interesting insights have emerged from a brain-machine interface experiment in which animals are trained to control a reward-delivering device by self-modulating the firing of a predefined set of neurons. Neurons causally involved in the task synchronized their firing around the up phase of slow waves during subsequent sleep epochs. Further, the performance improvement at retrieval could be predicted by this synchrony increase and was impaired by specific optogenetic silencing of activity during the up phase of the slow waves (*20*). Most cortical studies have focused on up states while largely ignoring silent phases. Indeed, the way we study the brain suffers from technical, statistical, and conceptual biases, and we tend to look at what we can most easily record and decode: periods of high population activity, higher firing neurons, and salient oscillatory patterns. An original approach both using and getting around these biases showed that the very sparse, usually dismissed activity in the prefrontal cortex during the prominent delta waves (down states) actually reactivated cell assemblies formed during preceding learning (*21*).

SWRs and cortical NREM sleep patterns are temporally coordinated in a manner that is believed to promote plasticity and long-term consolidation of contextual (or episodic) memories (*16*, *22*). The incidence of hippocampal SWRs is increased at transitions to cortical up and down

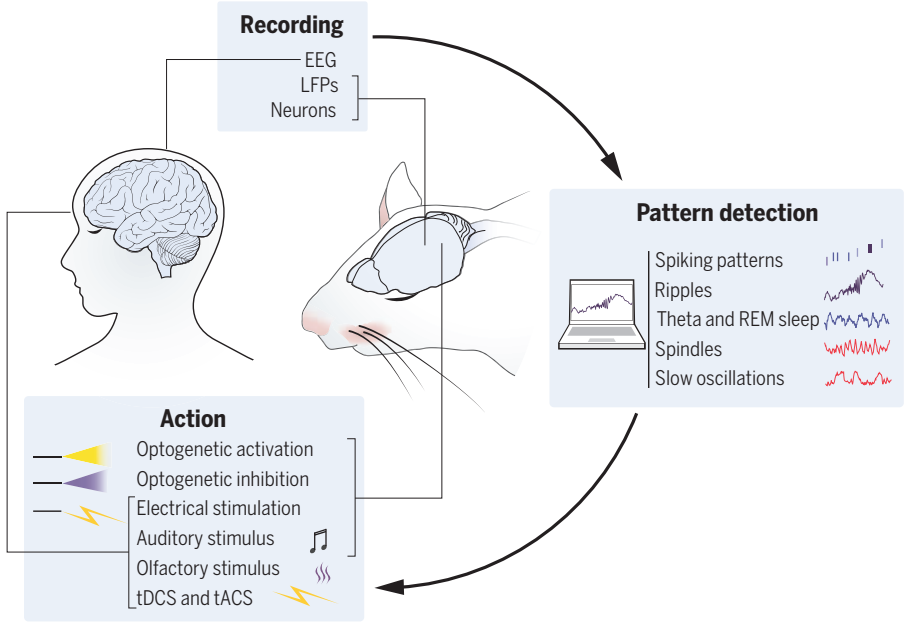


Fig. 2. Closed-loop experiments allow for the modulation of ongoing brain patterns in real time. Recorded brain signals are processed in real time to detect sleep patterns. The detection of a given pattern automatically triggers an action using invasive or noninvasive methods that affect the neural networks in real time to test whether the manipulation boosts or impairs memory consolidation. The effect on memory is assessed during a recall session after the modified sleep period. tDCS, transcranial direct current stimulation; tACS, transcranial alternating current stimulation.

states and spindle troughs, and coordinated reactivations occur between the hippocampus and various cortical areas during SWRs (*23*). Indeed, enhancing hippocampo-cortical coordination by using a closed-loop system (Fig. 3) to generate a down state–spindle complex after SWRs improves performance on a memory task (*17*). Optogenetically generating artificial spindles in coordination with hippocampal ripples and slow cortical oscillations also improves memory (*18*), highlighting the importance of the ripple-delta-spindle trifecta coordination for memory consolidation. Further, the spiking content of hippocampal SWRs can predict cortical firing in subsequent delta waves, suggesting that hippocampal SWRs bias the reactivated information in the cortex (*21*). Conversely, cortical firing can also predict the reactivated content in CA1 (*24*), and sensory stimulation during sleep can bias the content of hippocampal reactivation and improve memory, a phenomenon called targeted memory reactivation (*25*). Altogether, these findings indicate that memory consolidation involves loops where cortical areas can bias memory traces reactivated in hippocampal SWRs, which in turn would evoke the reactivation of related multimodal representations in the neocortex.

Beyond the hippocampo-cortical sleep talking

Because of the robust conceptual framework provided by both the two-step consolidation theory and the idea of a gradual transfer of information from the hippocampus toward cortical areas, most studies on sleep patterns and memory consolidation have focused on the hippocampo-cortical dialogue. However, many other structures are involved in memory formation. SWRs in the hippocampus are extremely powerful events that can synchronize activity across structures beyond the neocortex, potentially associating other features, such as emotional tone, to various forms of memories during the consolidation process. For example, reward-located hippocampal place cells and reward-encoding ventral striatum neurons fire together during sleep SWRs after the rewarded experience, with hippocampal activity leading the striatal activity (*26*). Dorsal versus ventral hippocampus SWRs modulate distinct populations of neurons in the nucleus accumbens (*27*), another crucial structure for reward processing. In the basolateral amygdala, a major center for valence encoding, a subset of neurons is modulated during hippocampal SWRs. The joint hippocampal-amygdala neuronal representation established during an aversive spatial experience is reinstated during the following NREM epoch, specifically during SWRs (*28*). These results suggest that hippocampal SWRs could be coordinators of brain-wide, plasticity-enabling activity or reactivation, allowing for the formation of distributed engrams across cortical, but also noncortical, areas.

REM sleep and theta oscillations

Despite long-standing general interest in REM sleep stemming from its association with vivid dreaming in humans, the functional physiology of REM sleep has been understudied compared with NREM sleep. REM sleep EEG and LFP activity closely resembles awake activity: It was originally called "paradoxical" sleep for this very reason. Indeed, the dominant rhythm during REM sleep is the theta oscillation, characteristic 5- to 12-Hz waves that are most prominent in the hippocampus but are recorded in the cortex and other subcortical structures as well. During wakefulness, hippocampal theta oscillations organizes place cell firing in sequences. This fine timing of hippocampal activity by theta oscillations is crucial for the encoding and subsequent consolidation of spatial memory through place cell replay during NREM sleep ripples (*29*). Comparatively, few studies have focused on how neuronal activity is structured during REM sleep, related to or independently of theta oscillations (*30*–*32*). Transient increases of theta frequency and power during REM, referred to as phasic REM, are associated with an increase in firing rate and coordination throughout the hippocampus and with cortical areas (*32*, *33*). Phasic REM has also been linked to the ponto-geniculo-occipital waves originating from the brainstem and has been suggested to coordinate various structures during REM sleep (*34*).

To date, the link between these specific changes in REM sleep theta dynamics and behavior remains unclear. However, the coherence between theta oscillations in the hippocampus, medial prefrontal cortex, and amygdala increases after aversive learning (*35*) in correlation with behavioral performance. The disruption of theta oscillations during REM sleep by optogenetically targeting the medial septum impairs hippocampus-dependent contextual memory consolidation (*36*). Additionally, the alteration of the activity of adult-born hippocampal neurons in the dentate gyrus specifically during REM sleep impaired contextual fear consolidation (*37*). Although the manipulations did not affect theta oscillations, the fact that both an increase or decrease of firing impaired consolidation suggests that the fine-timing—potentially theta-paced—firing of newborn neurons is important. Moreover, slight structural modifications of synapses in newborn neurons were reported upon REM-sleep inhibition, indicative of weakened synapses. These results add to previous studies that established that REM sleep promotes dendritic spine selective reinforcement or suppression in the neocortex (*38*). More work remains to be done to bridge the fine timing of patterned firing and theta oscillations during REM sleep with the observed structural plasticity in specific neuronal subpopulations and correlate it with behavioral outcomes.

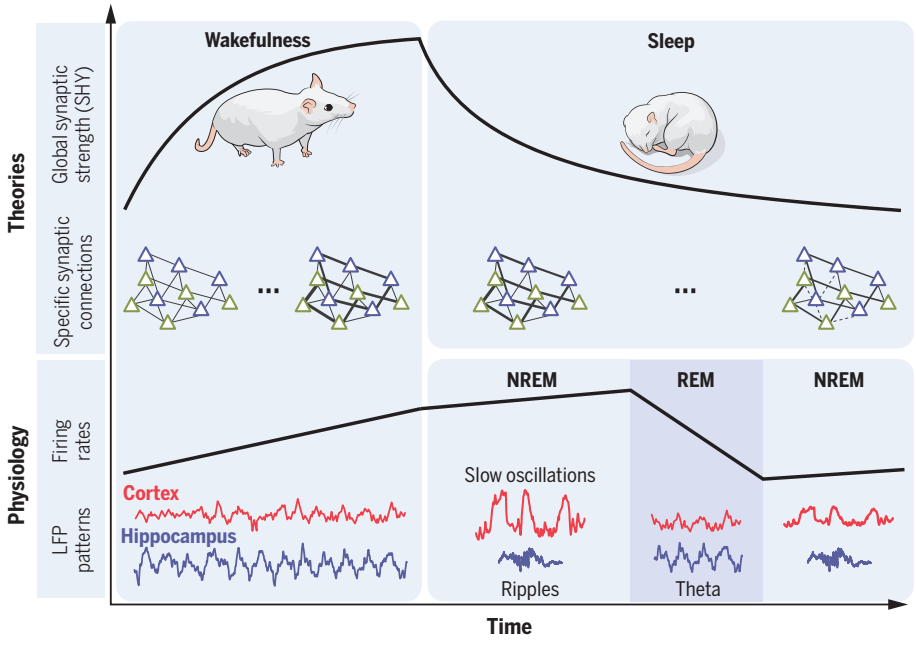


Fig. 3. Homeostasis and memory consolidation may occur in parallel across wake-sleep cycles. During learning, synapses are globally enhanced and overall firing rates progressively increase as the brain encodes new information into cell assemblies paced by theta oscillations in the hippocampus. (Engrams are shown as green and blue triangles; the thickness of the black line represents the strength of the connection.) During extended sleep periods, including sequences of NREM and REM epochs, homeostatic processes that involve cortical slow oscillations and REM sleep theta oscillations combine to downscale overall firing rates and global synaptic strength in accordance with SHY. In parallel, the specific connections among cell assemblies are selectively consolidated through ripple-related temporally organized reactivation (see Fig. 1).

Sleep rhythms and plasticity: Consolidation and homeostasis

Learning has been associated with Hebbian plasticity and synaptic potentiation. According to the synaptic homeostasis hypothesis (SHY), sleep plays a crucial role in homeostatic regulation by down-scaling synaptic weights to avoid saturation and allow for the formation of new memories during the subsequent wakefulness epoch. More specifically, this model predicts that global synaptic weights increase during wakefulness and decrease throughout sleep. Although there is structural and molecular evidence for this process (*39*), it is difficult to assess structural changes and strength in synapses in vivo and in real time. Because cortical slow-wave activity stems from highly synchronized activity through up and down states, their amplitude is thought to reflect synaptic strength between cortical neurons. Accordingly, slow oscillations are strongest after extended wakefulness and progressively diminish across prolonged sleep episodes, in line with the SHY model [Fig. 3; (*40*)]. Further, the changes in the slope of evoked potentials in the cortex, a marker of synaptic efficacy, are correlated with the changes in slow-wave activity, suggesting that slow waves might contribute to synaptic downscaling (*40*). In parallel, the dynamics of firing rates across awake and sleep periods have been used as a proxy for neuronal excitability. Coherent with the SHY model, hippocampal cells, as a population, progressively increase their firing rate during waking (*41*, *42*). During sleep, there is a global net decrease in firing rates but opposing trends between different stages: Whereas overall spiking activity increases during NREM, it shows a marked decrease during REM (*41*, *42*). Notably, downregulation of the firing rate during REM could be predicted by spindles and SWR incidence during NREM (*42*). Finally, the canonical NREM sleep SWRs, for the longest time thought to be propitious to consolidation through long-term potentiation (*43*), also trigger long-term depression (*44*), and their inhibition prevents the normal decrease of evoked potentials across sleep, suggesting a potential role in homeostasis.

Perspectives

Although simple to state, the link between sleep and memory actually translates into an incredibly complex field of research. First, sleep is not homogeneous and is subdivided into stages and substages that are characterized by different rhythms and patterns. Second, there are many different types of memories (episodic and semantic memories, procedural and skills memory, pavlovian conditioning, etc.) that rely on different, although sometimes overlapping, networks of structures, themselves exhibiting different sleep patterns. Further, episodic memories are not a complete and faithful representation of actual events. Episodic memory formation, therefore, encompasses the initial encoding of the information, modifications, merging with other memories, and even forgetting (*45*). Given the complexity of sleep, memory, and the diversity of the involved structures, how do we design relevant basic research unraveling "the role of sleep for memory"?

In rodents, NREM sleep is traditionally studied as a homogeneous stage. Characterizing more-specific NREM substages, or microstates, that potentially match the three human NREM substages is an interesting avenue to link them with various aspects of memory processing up to the behavioral level. The function of phasic versus tonic REM sleep in both humans and other species also remains to be investigated. In parallel, the study of patterns focuses on the function of specific network processes outside the frame of strictly defined stages. Indeed, several processes might coexist within a stage and could be more reliably identified by linking them to specific patterns rather than the stage as a whole. The development of closed-loop systems and brain-machine interfaces for real-time pattern detection in neuronal firing and EEG or LFP signals brought about major advancements in understanding the involvement of sleep patterns in memory formation (*8–10*, *13*, *14*, *17*, *36*). Sequences of place cells that represent experienced trajectories are reactivated in subsequent sleep SWR (*6*), but to date, there is no causal evidence that the sequence per se, as opposed to the mere activation of the place cell assembly (or engram) within a short time window, is important for memory consolidation. Testing theories on the importance of spike timing during patterns will require more-precise real-time tools to perturb or impose the precise timing relationships between specific neurons without altering their firing rate at a broader time scale (*46*). In turn, clarifying the question of the relevance of the sequence itself would potentially reorient the field toward the nearly 80% of SWRs for which the associated neuronal content cannot be identified as statistically significant sequences by the current decoding algorithms. These could be reactivation events that we are not yet capable of reading the way that downstream reader brain structures do or a replay of remote memories not assessed by the experimenter. According to this hypothesis, the main function of SWR-related high-synchronous events, including the ones we cannot decode, is to promote consolidation by means of memory replay. Another emerging and more integrated theory is that during sleep, the cortex and hippocampus enter default modes that result from their physiological properties and hardwiring, involving bouts of heightened and synchronized activity (SWRs and up states). These modes would primarily serve a homeostatic purpose (Fig. 3), but wakefulness activity and memory encoding would bias the precise timing of the firing during these events away from randomness, in which case specific memory traces could be consolidated (*44*, *47*). Further, the bias induced by wakefulness activity would be stronger and more long-lasting after learning or novelty (*11*), leading to periods of higher replay-to-noise ratio in SWR events. In that view, homeostasis and consolidation are on the same spectrum and heavily depend on the fine timing of the neuronal activity within the canonical sleep patterns.

"...during sleep, the cortex and hippocampus enter default modes that result from their physiological properties and hardwiring..."

Finally, reactivation, the main proposed mechanism for consolidation, is not universal in terms of structure and sleep stage, whereas homeostasis has been mostly studied in the neocortex. Therefore, more work needs to be done to precisely characterize sleep patterns in non–hippocampo-cortical structures that are involved in memory processing (e.g., amygdala, striatum). It is possible, and remains to be investigated, that consolidation and homeostatic processes differ or are absent in other structures, especially those with no detectable sleep reactivation and/or different firing-rate distributions across brain states. This direction is especially interesting for the highly complex network of structures that have a controlling role over sleep states and transitions such as the pons, thalamus, hypothalamus, locus coeruleus, and basal forebrain. Indeed, in the same way that consolidation and homeostasis might be tightly related, control and function of the different sleep stages could also be linked (*48*).

Fueled by emerging recording, manipulation, and analysis technologies with increasing spatiotemporal precision, we are in the process of completing a multidimensional knowledge space of mechanisms for different types of memory, different stages and substages of sleep, and their associated physiological patterns. Although we might never reach a unifying theory for the memory function of sleep, expanding and precising this space will allow us to better integrate consolidation and homeostasis, unravel new links within memory function in all steps of memory formation from encoding to retrieval through consolidation, and link mnemonic mechanisms with other aspects of sleep such as sleep control, circadian rhythm, or pathology.

REFERENCES AND NOTES

1. G. Buzsáki, *Hippocampus* **25**, 1073–1188 (2015).
2. G. Buzsáki, *Neuroscience* **31**, 551–570 (1989).
3. M. A. Wilson, B. L. McNaughton, *Science* **265**, 676–679 (1994).
4. V. Lopes-dos-Santos, S. Ribeiro, A. B. L. Tort, *J. Neurosci. Methods* **220**, 149–166 (2013).

5. D. Tingley, A. Peyrache, *Philos. Trans. R. Soc. London Ser. B* **375**, 20190231 (2020).
6. B. E. Pfeiffer, *Hippocampus* **30**, 6–18 (2020).
7. T. Schreiner, T. Staudigl, *Philos. Trans. R. Soc. London Ser. B* **375**, 20190293 (2020).
8. G. Girardeau, K. Benchenane, S. I. Wiener, G. Buzsáki, M. B. Zugaro, *Nat. Neurosci.* **12**, 1222–1223 (2009).
9. V. Ego-Stengel, M. A. Wilson, *Hippocampus* **20**, 1–10 (2010).
10. G. M. van de Ven, S. Trouche, C. G. McNamara, K. Allen, D. Dupret, *Neuron* **92**, 968–974 (2016).
11. B. Giri, H. Miyawaki, K. Mizuseki, S. Cheng, K. Diba, *J. Neurosci.* **39**, 866–875 (2019).
12. V. Lopes-Dos-Santos *et al.*, *Neuron* **100**, 940–952.e7 (2018).
13. A. Oliva, A. Fernández-Ruiz, F. Leroy, S. A. Siegelbaum, *Nature* **587**, 264–269 (2020).
14. I. Gridchyn, P. Schoenenberger, J. O'Neill, J. Csicsvari, *Neuron* **106**, 291–300.e6 (2020).
15. S. Lensu, T. Waselius, M. Penttonen, M. S. Nokia, *J. Neurophysiol.* **121**, 131–139 (2019).
16. J. G. Klinzing, N. Niethard, J. Born, *Nat. Neurosci.* **22**, 1598–1610 (2019).
17. N. Maingret, G. Girardeau, R. Todorova, M. Goutierre, M. Zugaro, *Nat. Neurosci.* **19**, 959–964 (2016).
18. C. V. Latchoumane, H.-V. V. Ngo, J. Born, H.-S. Shin, *Neuron* **95**, 424–435.e6 (2017).
19. L. Marshall, H. Helgadóttir, M. Mölle, J. Born, *Nature* **444**, 610–613 (2006).
20. T. Gulati, L. Guo, D. S. Ramanathan, A. Bodepudi, K. Ganguly, *Nat. Neurosci.* **20**, 1277–1284 (2017).
21. R. Todorova, M. Zugaro, *Science* **366**, 377–381 (2019).
22. I. Skelin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2012075118 (2021).
23. I. Skelin, S. Kilianski, B. L. McNaughton, *Neurobiol. Learn. Mem.* **160**, 21–31 (2019).
24. G. Rothschild, E. Eban, L. M. Frank, *Nat. Neurosci.* **20**, 251–259 (2017).
25. P. A. Lewis, D. Bendor, *Curr. Biol.* **29**, R906–R912 (2019).
26. C. S. Lansink, P. M. Goltstein, J. V. Lankelma, B. L. McNaughton, C. M. A. Pennartz, *PLOS Biol.* **7**, e1000173 (2009).
27. M. Sosa, H. R. Joo, L. M. Frank, *Neuron* **105**, 725–741.e8 (2020).
28. G. Girardeau, I. Inema, G. Buzsáki, *Nat. Neurosci.* **20**, 1634–1642 (2017).
29. C. Drieu, R. Todorova, M. Zugaro, *Science* **362**, 675–679 (2018).
30. K. Louie, M. A. Wilson, *Neuron* **29**, 145–156 (2001).
31. A. Peyrache, M. M. Lacroix, P. C. Petersen, G. Buzsáki, *Nat. Neurosci.* **18**, 569–575 (2015).
32. S. M. Montgomery, A. Sirota, G. Buzsáki, *J. Neurosci.* **28**, 6731–6741 (2008).
33. D. G. de Almeida-Filho *et al.*, *Sci. Rep.* **11**, 13078 (2021).
34. J. F. Ramirez-Villegas *et al.*, *Nature* **589**, 96–102 (2021).
35. D. Popa, S. Duvarci, A. T. Popescu, C. Léna, D. Paré, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6516–6519 (2010).
36. R. Boyce, S. D. Glasgow, S. Williams, A. Adamantidis, *Science* **352**, 812–816 (2016).
37. D. Kumar *et al.*, *Neuron* **107**, 552–565.e10 (2020).
38. W. Li, L. Ma, G. Yang, W.-B. Gan, *Nat. Neurosci.* **20**, 427–437 (2017).
39. L. de Vivo *et al.*, *Science* **355**, 507–510 (2017).
40. V. V. Vyazovskiy *et al.*, *Neuron* **63**, 865–878 (2009).
41. A. D. Grosmark, K. Mizuseki, E. Pastalkova, K. Diba, G. Buzsáki, *Neuron* **75**, 1001–1007 (2012).
42. H. Miyawaki, K. Diba, *Curr. Biol.* **26**, 893–902 (2016).
43. J. H. L. P. Sadowski, M. W. Jones, J. R. Mellor, *Cell Rep.* **14**, 1916–1929 (2016).
44. H. Norimoto *et al.*, *Science* **359**, 1524–1527 (2018).
45. G. R. Poe, *J. Neurosci.* **37**, 464–473 (2017).
46. Z. Zhang, L. E. Russell, A. M. Packer, O. M. Gauld, M. Häusser, *Nat. Methods* **15**, 1037–1040 (2018).
47. D. Levenstein, B. O. Watson, J. Rinzel, G. Buzsáki, *Curr. Opin. Neurobiol.* **44**, 34–42 (2017).
48. S. Izawa *et al.*, *Science* **365**, 1308–1313 (2019).

ACKNOWLEDGMENTS

We thank V. Cazares, L. Roux, D. Almeida-Filho, and members of the Girardeau lab for helpful comments on the manuscript. We apologize to all authors whose studies we could not cite because of a focus on the most recent literature and space restrictions. We acknowledge that this selection of references does not reflect the highly incremental and collaborative nature of scientific research. **Funding:** This work was supported by an ATIP-Avenir grant; the Fyssen Foundation; Emergence(s) Ville de Paris; a NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation and the Schlumberger Foundation for Science and Education (G.G.); and the Medical Research Council UK (MC_UU_00003/4; V.L.d.S.). **Competing interests:** The authors declare no competing interests. This research was funded in whole or in part by the Medical Research Council UK (MC_UU_00003/4), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

10.1126/science.abi8370

REVIEW

The interconnected causes and consequences of sleep in the brain

Laura D. Lewis*

Sleep is essential for brain function in a surprisingly diverse set of ways. In the short term, lack of sleep leads to impaired memory and attention; in the longer term, it produces neurological dysfunction or even death. I discuss recent advances in understanding how sleep maintains the physiological health of the brain through interconnected systems of neuronal activity and fluid flow. The neural dynamics that appear during sleep are intrinsically coupled to its consequences for blood flow, cerebrospinal fluid dynamics, and waste clearance. Recognizing these linked causes and consequences of sleep has shed new light on why sleep is important for such disparate aspects of brain function.

Our nightly sleep is critical for a wide array of brain functions. Missing just a single night of sleep results in memory, mood, and attentional impairments the next day (*1*); disrupted sleep across the life span is linked to neurodegeneration (*2*, *3*). These heterogeneous effects of sleep have posed a puzzle in neuroscience: Why does this brain state hold a unique role in supporting such seemingly distinct aspects of brain function?

Decades of work have shown that sleep contains distinct neural dynamics linked to cognition, such as slow waves in neural activity that appear in non–rapid eye movement (NREM) sleep (*4*). More recent discoveries have shown that sleep is also a heightened state for waste removal from the brain (*5*, *6*). Metabolic waste products are transported out of brain tissue (*7*) via the interstitial fluid (ISF) and cerebrospinal fluid (CSF), and sleep plays a key role in both waste regulation and flow of CSF in the brain (*8*, *9*), which is essential for maintaining neuronal health. These beneficial effects may explain why we spend so many hours each day asleep, as sleep's role in basic housekeeping for the brain implicates it in broad aspects of neural function. But why is sleep linked to distinct fluid dynamics, and why does it play such a profound role in maintaining brain function?

The control circuits that govern sleep

Sleep is controlled by large-scale arousal regulatory systems that can rapidly induce states of sleep or wakefulness. Extensive circuit-based investigations have recorded from and manipulated individual brain regions and cell types to identify their various roles in behavior. This approach has been powerful, identifying multiple regions in the hypothalamus, brainstem, basal forebrain, and other subcortical nuclei that control sleep states (*10*). Key interacting circuits include the noradrenergic, dopaminergic, cholinergic, and orexinergic systems, which can induce sleep-wake transitions (*10*). However, these findings have raised a new question: Why do so many diverse nuclei play decisive roles in whether an animal is sleeping or awake?

One possibility is that the system is simply unusually redundant; sleep is so important that the brain contains multiple switches to induce sleep states. However, a second and perhaps more likely possibility is that these circuits represent interacting components of a larger system that induces the many distinct features of sleep, including altered behavior, oscillatory neural dynamics, modulated respiratory and vascular physiology, and clearance. Notably, the circuits that regulate sleep project widely throughout the forebrain (*10*) as well as to each other, so that activity within one system inevitably influences the others. Furthermore, neuromodulators such as noradrenaline and acetylcholine, which alter neuronal arousal states, also have direct actions on the vasculature.

Low-frequency neural dynamics emerge during sleep

During sleep, large-scale neural activity reorganizes into distinctive, oscillatory patterns, reflecting the wide-ranging reach of arousal regulatory circuits throughout the thalamocortical networks that generate EEG oscillations. These patterns have long been recognized to define distinct stages of sleep, separating NREM and REM. As sleep progresses into NREM, many distinct EEG patterns appear, including spindles (~11 to 14 Hz) and slow waves, with distinct links to memory and cognition (*4*). The increased low-frequency power is referred to as slow wave activity, a catch-all term that corresponds to 0.5 to 4 Hz EEG power, and which can reflect several low-frequency dynamics (Fig. 1A). These include individual K-complexes (occurring seconds or minutes apart), delta (1 to 4 Hz) waves, and the slow (0.1 to 1 Hz) oscillations that group and coordinate higher-frequency rhythms (*11*, *12*). K-complexes correspond to periods

Department of Biomedical Engineering, Boston University, Boston, MA, USA, and Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA, USA.
***Corresponding author. Email: ldlewis@bu.edu**

of widespread suppression of neural activity lasting hundreds of milliseconds (*13*) ("down-state"), typically throughout large expanses of the cortex (*14*). In deeper NREM sleep, slow waves become continuous and rhythmic, alternating between down-states and up-states. REM sleep, by contrast, is linked to desynchronized EEG states, as well as patterns such as rapid eye movements and suppressed muscle tone (*15*).

Although ample evidence demonstrates that brainwide activity changes during sleep and that slow waves can be global cortical events, this should not be taken to imply that sleep is a uniform, homogeneous state throughout the brain. Sleeplike slow waves can appear in local patches of the cortex even within the awake brain (Fig. 1B), termed "local sleep" (*16*). One possibility is that the brain's need for slow waves is so strong that after lack of sleep, slow waves emerge in the awake brain despite their detrimental consequences for behavior. During NREM sleep, slow waves originally thought to represent globally coherent activity have also been found to exhibit local dynamics (*17*). Slow wave activity is not nearly as prominent during REM sleep but can occur in superficial cortical layers or frontal regions (*18*).

Waste clearance and CSF flow in the sleeping brain

Decades of study have contributed to our understanding of the slow wave phenomena discussed above. More recent was the discovery of sleep's role in waste clearance: Molecules such as amyloid-β are cleared from the mouse brain at far higher rates during sleep than during wakefulness (*5*). This solute clearance takes place via ISF and CSF flowing along blood vessels (*7*); however, the precise mechanisms remain a topic of controversy, with debate over the forces that drive flow and the exit routes for solutes (*19*–*21*). This observation yielded a new perspective on the importance of sleep: Sleep maintains the basic physiological health of neurons by removing their potentially harmful metabolic waste.

An important consideration is that waste production rates also differ across arousal states, with higher tau and amyloid production during wakefulness in both rodents and humans (*8*). Sleep may thus serve as a pause in the waste generation process, allowing the clearance system time to catch up with the detritus that accumulates during wakefulness. The relative balance of these two processes—a respite from waste production, or a period of enhanced cleansing—needs further study.

Human imaging studies have provided recent support for the link between sleep and brain waste regulation. Sleep deprivation increases amyloid-β in the brains of healthy young adults (*22*). Furthermore, injections of a contrast agent revealed that clearance from brain tissue is higher when participants sleep than when they are kept awake (*6*). This impaired waste clearance is apparent after a single night of sleep deprivation—a striking observation, given that this is a not-infrequent behavior for many individuals.

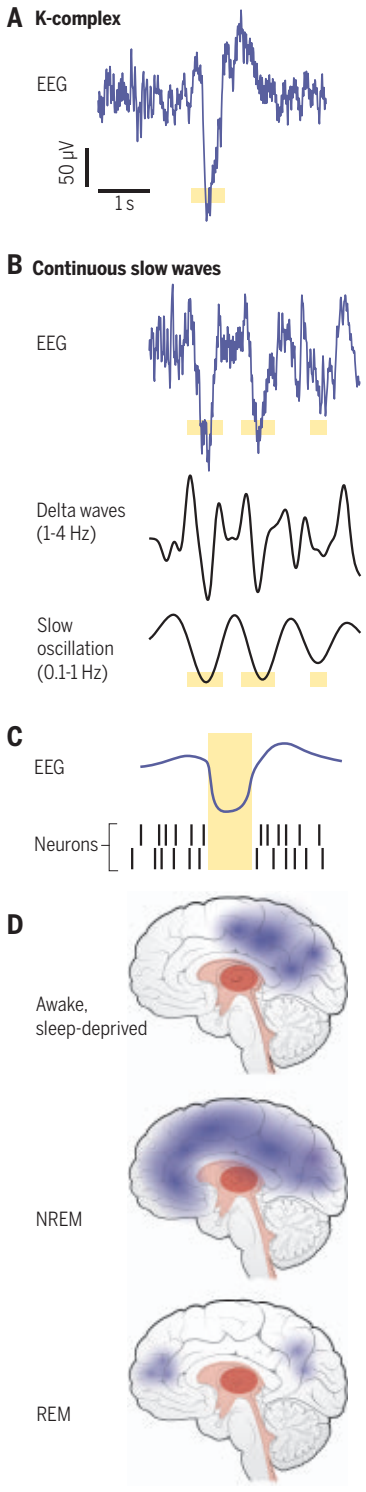


Fig. 1. Low-frequency neural activity during sleep. (**A** and **B**) Slow wave activity (0.5 to 4 Hz) appears in the EEG during NREM sleep, with emergence of individual K-complexes (A) or continuous slow waves (B). Yellow highlights denote possible down-states. (**C**) Illustration of slow waves coupled to neuronal suppression (yellow). [Adapted from (*13*, *17*)] (**D**) Low-frequency neural dynamics can occur locally in the awake brain, locally or globally in NREM, and in restricted cortical regions in REM.

Why does sleep increase brain clearance? One contributing factor is that extracellular volume expands during sleep (*5*), which would increase the rate of molecular transport. Second, higher clearance rates occur in rodents when using anesthetics that induce high delta power (*23*), hinting that the neural dynamics of sleep are related to clearance. Another factor is that fluid flow patterns change during sleep. It has long been known that CSF flow in the awake human brain constantly pulses with the cardiac and respiratory cycles (*24*), but CSF flow during sleep has only recently been investigated.

A recent imaging study in humans repurposed classic flow-related enhancement signals in functional magnetic resonance imaging (fMRI) to simultaneously measure EEG, blood oxygenation, and CSF flow during sleep (*9*). This imaging revealed large waves of CSF flow during NREM sleep. The CSF waves were preceded by neural slow wave activity several seconds earlier and were anticorrelated with hemodynamic signals. This temporal coupling was consistent with a model in which neural activity drives CSF flow through its effects on blood volume, which in turn displaces CSF (Fig. 2A). This mechanism would explain how the intrinsic neural dynamics of sleep are linked to fluid flow.

Neurovascular physiology contributing to CSF flow

What specific vascular mechanisms might implement this observed coupling of neural slow waves and CSF flow? This question highlights a continued challenge throughout sleep research: Many features of brain physiology undergo correlated changes. Specifically, neural slow waves are linked to glial activity, cognitive processes, autonomic state, and vascular dynamics. Several of these processes likely contribute to coupled neural and CSF flow waves. First, neural activity elicits local changes in blood volume through neurovascular coupling (*25*); this relationship is the basis of most fMRI studies. Low-frequency (~0.1 Hz) modulation of neural activity can also entrain arteriolar vasomotion, leading to fluctuations in blood volume (*26*). Because EEG slow waves correspond to widespread suppression of cortical firing, these neuronal changes would cause decreased blood volume and increased CSF flow. Furthermore, neurovascular coupling is strengthened during NREM sleep (*27*), bolstering this mechanism.

Second, slow waves during sleep not only reflect local neural activity but also are often coupled to systemic changes in vasoconstriction caused by altered neuromodulatory and autonomic states, particularly when slow waves

occur as isolated events termed K-complexes (*28*). This systemic vasoconstriction is also coupled to brainwide hemodynamic changes and CSF flow in humans (*29*, *30*). In addition, individual slow waves are phase-locked to slow oscillations (<1 Hz) (*11*), which are linked to autonomic modulation of blood flow (*12*). Indeed, this pathway may not be fully separable from neurovascular coupling; the systemic vasoconstriction may partially reflect the need for brainwide hemodynamic modulation in concert with the large changes in neuronal activity. The neuromodulators linked to sleep can also have direct effects on vasodilation; for example, the noradrenergic system modulates both sleep slow oscillations and vessel diameter (*31*, *32*). Its effects likely depend on whether release is tonic or phasic, which differs across sleep stages and would affect subsequent CSF flow. Low-frequency vascular oscillations (~0.1 Hz, i.e., a cycle every ~10 s) are also present at lower levels during wakefulness. Consistent with the vascular-based model (Fig. 2), these low-frequency vascular dynamics are also coupled to CSF flow during wakefulness, but with lower amplitude than during sleep (*33*). Similarly, modulating breathing (which affects vasodilation) affects CSF flow during wakefulness (*34*). Thus, in this framework, the temporal properties of the vasculature are the key component that sets the timing of fluid flow, and coherent low-frequency neural activity during sleep is therefore a particularly effective driver of flow.

A key prediction of the model proposed here is that neural activity that is the most effective at entraining vascular changes will drive the largest CSF flow (Fig. 2B). Low-frequency EEG dynamics during sleep display many distinct patterns (Fig. 1); this model predicts that any of these slow dynamics could drive flow if it is linked to widespread changes in vascular dilation. For example, an isolated slow wave would be predicted to drive CSF flow more effectively than continuous slow waves because of the slow vascular response (Fig. 2B). In addition, lower-frequency oscillations (e.g., slow oscillations) are predicted to be more effective than higher frequencies (e.g., delta waves). Furthermore, slow oscillations are phase-coupled to the amplitude of higher-frequency dynamics such as spindles. Studies of total sleep deprivation cannot disambiguate between the roles of distinct neural rhythms, and further study is needed to test whether different oscillations have different links to flow. Given that vascular mechanics may be the critical element governing fluid flow and clearance, it can likely be induced by multiple types of coherent neural activity or slow modulators of vasodilation. Recent work in mice supports this idea: Presentation of low-frequency (0.05 Hz, or every 20 s) visual stimuli entrained arteriole dilation and enhanced paravascular clearance (*35*).

Neural and fluid dynamics at the mesoscale

A key open challenge is to bridge the macroscopic and microscopic scales of fluid dynamics observed in sleep. Rodent studies established clearance rates by monitoring solute transport along vessels (*7*, *21*). Human sleep studies have observed macroscopic CSF flow in the ventricle (*9*) and brainwide protein accumulation (*6*). A major question concerns how these scales are linked: How does the bulk flow of CSF in the ventricles affect clearance in the tissue, and is neurovascular coupling a viable mechanism for driving solute transport out of the brain? Although experimental access at the mesoscale is challenging, computational models have shed light on these questions. Models show that the slow time scale and large amplitude of neurovascular coupling make it an effective mechanism for driving solute transport along arterioles (*36*, *37*).

Studies measuring macroscopic CSF flow in humans have not yet specifically linked CSF velocity in the ventricles to rates of clearance. Intuitively, the idea of high-velocity CSF flow waves during sleep might be expected to increase clearance, analogous to how a stagnant bath differs from one where the water is constantly mixed and refreshed. However, this intuition has not yet been empirically confirmed, and future studies will be needed to determine the precise relationship between large-scale CSF flow and solute transport out of brain tissue.

Closing the loop: The consequences of fluid physiology for neuronal function

An emerging and tantalizing question is whether the effect of neural activity on CSF flow forms part of a bidirectional feedback loop, where each can influence the other. Several studies suggest specific routes by which fluid contents might modulate arousal. The ionic composition of the ISF can modulate neuronal firing and induce states of wakefulness or sleep (*38*). Amyloid and inflammatory cytokines also affect neural arousal state (*39*, *40*). By modulating the local milieu of molecular composition of the ISF and CSF, clearance may thus affect sleep. Furthermore, individuals with genotypes linked to low expression of aquaporin-4, which forms part of the glymphatic pathway (*7*), show higher EEG slow wave activity (*41*). This observation has been construed as EEG slow wave activity

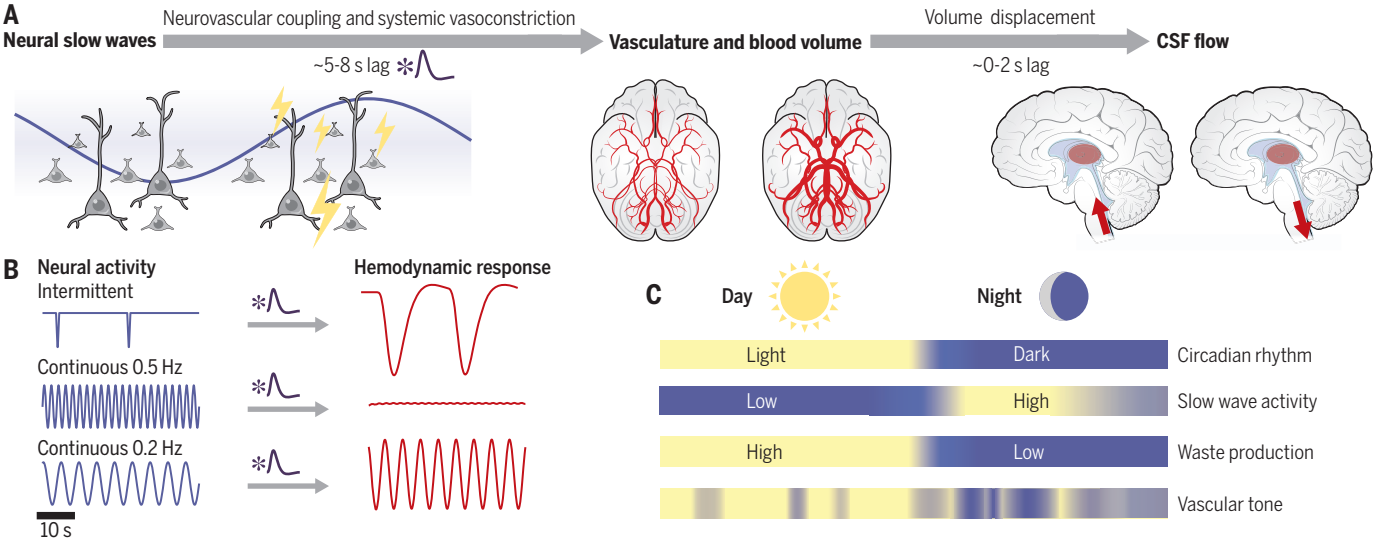


Fig. 2. Disparate time scales of coupled neural, vascular, and CSF dynamics in sleep. (**A**) EEG slow wave activity (0.5 to 4 Hz) reflects coherent changes in neural activity. Slow coherent neural activity is linked to slow dilation and constriction of blood vessels. These blood volume changes drive CSF flow. (**B**) Classic hemodynamic models produce slow temporal dynamics and predict that infrequent or slow neural activity elicits larger responses. (**C**) Multiple factors contribute on distinct time scales, including circadian rhythms and fluctuations in autonomic state.

increasing to compensate for lower clearance rates, although how such compensatory feedback might be implemented is unknown.

Moreover, the longer-term consequences of clearance—or lack thereof—may have more fundamental consequences for neuronal health, leading to inflammation or neurodegeneration in regions lacking adequate clearance. If clearance impairments affect the arousal regulatory circuits that induce sleep, this would further reduce sleep (*42*). Furthermore, as noted above, wakefulness is linked not only to lower clearance but also to higher rates of waste production in the first place (*8*). This "vicious cycle" hypothesis could explain why disrupted sleep is linked to the development of neurodegenerative disorders.

The linked neural and vascular systems also point to a dual vulnerability in aging. The length and depth of sleep both decline in aging. Although some decrease in sleep is typical of healthy aging, greater sleep loss is predictive of subsequent Alzheimer's disease pathology (*2*, *3*). EEG slow waves in particular are implicated: Patients with reduced slow wave activity show lower memory scores and increased gray matter atrophy (*43*). Decreased slow waves (<1 Hz) during sleep predict the subsequent accumulation of amyloid years later (*44*). Although causal evidence is yet to be established, the link between EEG slow waves and fluid dynamics suggests that loss of slow waves, particularly in the lowest-frequency bands, could impair clearance.

In addition to the declining neural signatures of sleep in aging, neurovascular physiology is also disrupted. Vascular dysfunction may be an early trigger for Alzheimer's disease, as cerebral blood flow declines years before symptom onset (*45*); this decline may also result in a failure of sleep to drive effective clearance, because vascular dilations drive CSF flow. In support of this idea, CSF flow imaging (*9*) was recently applied to a database from patients with mild cognitive impairment. Intriguingly, patients with such impairment showed weaker coupling between hemodynamics and CSF flow (*46*), which suggests that indeed this vascular mechanism for driving CSF flow may be impaired at early stages of neurodegeneration.

Fluid physiology across sleep stages

Finally, it is unclear how distinct sleep stages differ in their contribution to fluid flow. Certainly, the mechanism discussed here addresses one part of the question, identifying how neural and fluid dynamics are linked in NREM sleep, particularly in stages such as N2 sleep where slow waves are apparent but irregularly timed. Humans spend most of their NREM sleep in stage N2 (*47*), and the infrequent slow wave timing in this state may be particularly effective at driving CSF flow because of the slow vascular filter (Fig. 2B). However, local slow waves and changes in systemic vasodilation can occur even within the awake brain; consequently, the sleeplike CSF flow waves may also occur less frequently during wakefulness and lighter N1 sleep. Complicating these questions, substages of NREM in rodents are not clearly mapped to human sleep stages, and the timing of neural oscillations and neurovascular coupling can differ. Although these species-specific differences can pose challenges in translation, they also present an opportunity for dissecting how distinct neural and physiological dynamics modulate clearance systems within NREM.

In contrast, how clearance and CSF dynamics change during REM sleep is still unknown. Brainwide hyperemic patterns have been recently observed in rodents during REM sleep (*48*), and arterioles demonstrate unusually large dilations (*27*). These large fluctuations in blood volume may also drive CSF flow, but this possibility has not yet been tested. Because slow waves are not prominent in REM sleep, a distinct mechanism may be driving fluid dynamics during REM. One possibility is direct action of the REM-linked neuromodulators on the vasculature; for example, the cholinergic system, which is highly active in REM, can also have direct vasodilatory effects. Furthermore, circadian cycles also affect clearance (Fig. 2C) (*49*, *50*). Neural activity alone thus cannot explain the full picture of clearance during sleep, and more work is necessary to understand how these mechanisms act together.

Outlook and open questions

Sleep has diverse effects on the brain; neural activity and cognition are transformed, systemic and autonomic physiology shifts, and critical housekeeping processes support neuronal health. These processes have often been studied separately, but they are intrinsically linked through their mechanistic origins and physiological consequences. In turn, the neural dynamics that appear during sleep shape vascular and CSF flow, which feed back into these neural dynamics.

These converging results point to key open questions. First, what are the neural circuits that control clearance and fluid flow during sleep? Despite the work outlined above, which highlights a role for neural activity, how the diversity of specific neural signatures and sleep stages modulates fluid dynamics is not well understood. Given the astonishing number of neural circuit pathways that have been shown to control sleep, future work should identify how their interactions shape not only neural activity but also vascular dynamics, CSF flow, and clearance.

Recently developed technologies are increasingly making such brainwide, multimodal imaging studies possible. The ability to record at large scales in animal models enables studies of the joint, spontaneous dynamics across the circuits that shape sleep. In human neuroscience, the impressive recent advances in the spatiotemporal resolution of fMRI also place many of these questions within reach. With these new technologies, the field is poised to make major advances toward learning how these distributed dynamics interact to produce sleep states.

A second key challenge is to understand the mechanistic links between these interacting neuronal and non-neuronal systems, which pose a challenge for experimental investigation. Given that so many features of sleep are strongly correlated, dissecting causal relationships is difficult, and many conventional approaches include assumptions that preclude discovering these links. fMRI studies often will simply regress out one feature, such as respiration, but this presumes that these systemic physiological dynamics are a purely confounding factor; in fact, during sleep, neural state is often collinear with and drives systemic physiology. Systems neuroscience approaches often manipulate one circuit to make statements about causality but can miss the cascade of subsequent activity that results from focal manipulations, and directly modulating neural activity can sometimes produce effects unlike those that occur spontaneously. In addition to the neural and fluid dynamics outlined here, sleep also serves diverse other functions for the brain, such as synaptic homeostasis, glial function, memory, and dreaming. Considering brain physiology during sleep as an interconnected dynamic system, via multimodal studies that simultaneously capture distinct aspects of sleep, is a promising perspective for understanding how these processes interact.

The results of such studies will be critical for interpreting how sleep is linked to neurological and psychiatric disorders. The decline of sleep in neurodegenerative disorders is now clearly established, highlighting the need to identify the precise consequences of sleep for brain health. The relationship between clearance and psychiatric disorders is a much less explored area that merits further study, because disordered sleep is a signature of several psychiatric disorders (*51*). Achieving a mechanistic understanding is needed not just to probe and predict neural function linked to sleep, but to identify targets that may enable sleep-based interventions for improving brain health and clinical outcomes.

Ultimately, many factors act together to produce the effects of sleep, including coherent neural activity, vascular dynamics, and CSF and ISF flow. Although these interacting components can make it challenging to probe individual mechanisms experimentally, considering these dynamics as a whole can reveal their biophysical links. Sleep's powerful modulation of these many interconnected brain systems underlies its diverse and wide-ranging effects in maintaining cognition and healthy brain function.

REFERENCES AND NOTES

1. A. J. Krause *et al.*, *Nat. Rev. Neurosci.* **18**, 404–418 (2017).
2. B. P. Lucey *et al.*, *Sci. Transl. Med.* **11**, eaau6550 (2019).
3. S. Sabia *et al.*, *Nat. Commun.* **12**, 2289 (2021).
4. J. G. Klinzing, N. Niethard, J. Born, *Nat. Neurosci.* **22**, 1598–1610 (2019).
5. L. Xie *et al.*, *Science* **342**, 373–377 (2013).
6. P. K. Eide, V. Vinje, A. H. Pripp, K.-A. Mardal, G. Ringstad, *Brain* **144**, 863–874 (2021).
7. J. J. Iliff *et al.*, *Sci. Transl. Med.* **4**, 147ra111 (2012).
8. J. K. Holth *et al.*, *Science* **363**, 880–884 (2019).
9. N. E. Fultz *et al.*, *Science* **366**, 628–631 (2019).
10. F. Weber, Y. Dan, *Nature* **538**, 51–59 (2016).
11. F. Amzica, M. Steriade, *Neurology* **49**, 952–959 (1997).
12. P. Achermann, A. A. Borbély, *Neuroscience* **81**, 213–222 (1997).
13. S. S. Cash *et al.*, *Science* **324**, 1084–1087 (2009).
14. L. D. Lewis *et al.*, *eLife* **7**, e33250 (2018).
15. C. Iber, S. Ancoli-Israel, A. L. Chesson, S. F. Quan, *The AASM Manual for the Scoring of Sleep and Associated Events* (American Academy of Sleep Medicine, Westchester, IL, ed. 1, 2007).
16. V. V. Vyazovskiy *et al.*, *Nature* **472**, 443–447 (2011).
17. Y. Nir *et al.*, *Neuron* **70**, 153–169 (2011).
18. C. M. Funk, S. Honjoh, A. V. Rodriguez, C. Cirelli, G. Tononi, *Curr. Biol.* **26**, 396–403 (2016).
19. E. N. T. P. Bakker, D. M. P. Naessens, E. VanBavel, *Exp. Physiol.* **104**, 1013–1017 (2019).
20. H. Mestre, Y. Mori, M. Nedergaard, *Trends Neurosci.* **43**, 458–466 (2020).
21. N. Agarwal, R. O. Carare, *Front. Neurol.* **11**, 611485 (2021).
22. E. Shokri-Kojori *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4483–4488 (2018).
23. L. M. Hablitz *et al.*, *Sci. Adv.* **5**, eaav5447 (2019).
24. D. A. Feinberg, A. S. Mark, *Radiology* **163**, 793–799 (1987).
25. C. Iadecola, *Neuron* **96**, 17–42 (2017).
26. C. Mateo, P. M. Knutsen, P. S. Tsai, A. Y. Shih, D. Kleinfeld, *Neuron* **96**, 936–948.e3 (2017).
27. K. L. Turner, K. W. Gheres, E. A. Proctor, P. J. Drew, *eLife* **9**, e62071 (2020).
28. I. M. Colrain, *Sleep* **28**, 255–273 (2005).
29. P. S. Özbay *et al.*, *Commun. Biol.* **2**, 421 (2019).
30. D. Picchioni *et al.*, bioRxiv 442672 [preprint]. 5 May 2021.
31. C. W. Berridge, B. E. Schmeichel, R. A. España, *Sleep Med. Rev.* **16**, 187–197 (2012).
32. C. M. Peppiatt, C. Howarth, P. Mobbs, D. Attwell, *Nature* **443**, 700–704 (2006).
33. H.-C. Yang *et al.*, bioRxiv 437406 [preprint]. 9 September 2021.
34. S. Dreha-Kulaczewski *et al.*, *J. Neurosci.* **35**, 2485–2491 (2015).
35. S. J. van Veluw *et al.*, *Neuron* **105**, 549–561.e5 (2020).
36. A. K. Diem, R. O. Carare, R. O. Weller, N. W. Bressloff, *PLOS ONE* **13**, e0205276 (2018).
37. R. T. Kedarasetti *et al.*, *Fluids Barriers CNS* **17**, 52 (2020).
38. F. Ding *et al.*, *Science* **352**, 550–555 (2016).
39. G. G. Özcan, S. Lim, P. Leighton, W. T. Allison, J. Rihel, *eLife* **9**, e53995 (2020).
40. L. Besedovsky, T. Lange, M. Haack, *Physiol. Rev.* **99**, 1325–1380 (2019).
41. S. M. Ulv Larsen *et al.*, *PLOS Biol.* **18**, e3000623 (2020).
42. C. Wang, D. M. Holtzman, *Neuropsychopharmacology* **45**, 104–120 (2020).
43. B. A. Mander *et al.*, *Nat. Neurosci.* **16**, 357–364 (2013).
44. J. R. Winer *et al.*, *Curr. Biol.* **30**, 4291–4298.e3 (2020).
45. K. Kisler, A. R. Nelson, A. Montagne, B. V. Zlokovic, *Nat. Rev. Neurosci.* **18**, 419–434 (2017).
46. F. Han *et al.*, *PLOS Biol.* **19**, e3001233 (2021).
47. M. M. Ohayon, M. A. Carskadon, C. Guilleminault, M. V. Vitiello, *Sleep* **27**, 1255–1273 (2004).
48. A. Bergel, T. Deffieux, C. Demené, M. Tanter, I. Cohen, *Nat. Commun.* **9**, 5364 (2018).
49. X. Cai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 668–676 (2020).
50. L. M. Hablitz *et al.*, *Nat. Commun.* **11**, 4411 (2020).
51. R. M. Benca, W. H. Obermeyer, R. A. Thisted, J. C. Gillin, *Arch. Gen. Psychiatry* **49**, 651–668 (1992).

ACKNOWLEDGMENTS

I thank M. Nedergaard, R. Stickgold, and J. Voigts for discussions. **Funding:** NIH grants R01-AG070135 and R01-AT011429, 1907 Trailblazer Award from 1907 Foundation, NARSAD Young Investigator Award from the Brain and Behavior Research Foundation, Searle Scholars Program, One Mind Rising Star Award, and Simons Collaboration on Plasticity in the Aging Brain (no. 811231). **Competing interests:** The author is a co-inventor on patent application PCT/US20/27963, "Generating Imaging-Based Neurological State Biomarkers and Estimating Cerebrospinal Fluid (CSF) Dynamics Based on Coupled Neural and CSF Oscillations During Sleep," and has received speaking fees from Roche.

10.1126/science.abi8375

REVIEW

The translational neuroscience of sleep: A contextual framework

Michael A. Grandner[1]* and Fabian-Xosé Fernandez[2]

Sleep is entwined across many physiologic processes in the brain and periphery, thereby exerting tremendous influence on our well-being. Yet sleep exists in a social-environmental context. Contextualizing sleep health with respect to its determinants—from individual- to societal-level factors—would enable neuroscientists to more effectively translate sleep health into clinical practice. Key challenges and opportunities pertain to (i) recognizing and exploring sleep's functional roles, (ii) clarifying causal mechanisms in relation to key outcomes, (iii) developing richer model systems, (iv) linking models to known contextual factors, and (v) leveraging advances in multisensory technology. Meeting these challenges and opportunities would help transcend disciplinary boundaries such that social-environmental considerations related to sleep would become an ever-greater presence in the clinic.

Sleep is generally defined as a naturally recurring, reversible state of perceptual disengagement, reduced consciousness, and relative immobility, the propensity of which is patterned by homeostatic and circadian factors. Many studies have described the neurophysiology that contributes to sleep regulation, including the neuroanatomy of sleep and arousal-promoting circuitry (*1*); the diffuse, interconnected neurotransmitter systems that transition between non–rapid eye movement sleep (NREM), REM sleep, and wakefulness (*2*); and the core molecular clock mechanism and circadian pacemaker activity within the brain (*3*). Additional work has delineated how sleep serves as a prerequisite for optimizing the performance of critical biological functions, including those related to memory (*4*), waste clearance from the brain (*5*), and nutrient metabolism (*6*). Over the past 70 years, sleep has been increasingly conceptualized with deference to these findings, promoting an understanding of sleep as a set of interrelated physiological processes sometimes devoid of greater biological and social-environmental context.

Sleep is not just a collection of physiologic processes that occur together at the intersection of rest and activity and of circadian light-dark rhythms. Sleep is a non-negotiable biological state required for the maintenance of human life. It is associated with behaviors, but sleep itself is not a volitional behavior. It involves physiologic processes but is more than an amalgam of them. Sleep is a foundational aspect of our biology—our need for sleep parallels those for air, food, and water.

"Sleep is a non-negotiable biological state required for the maintenance of human life. ... our need for sleep parallels those for air, food, and water."

Total sleep deprivation results in death in animal models (*7*). For obvious ethical reasons, no previous scientific work has empirically tested the amount of sleep deprivation necessary to induce organ failure and death in humans. Prolonged sleep deprivation in humans has been studied, though. When sleep is curtailed, sleep propensity increases, such that individuals kept from sleep for extended periods will fall asleep faster and more deeply when sleep is no longer restricted (*8*). Continued wakefulness for several days results in the occurrence of involuntary sleep, where individuals fall asleep even while trying to stay awake. Although there are no formal studies of human mortality related to in-lab sleep loss, there are many reported cases in which sleep propensity was sufficiently high to induce involuntary sleep initiation, even during activities when such an event would likely result in death (such as while driving a car) (*9*). Thus, sleep is not a luxury for people with too much free time; it is a biological necessity whose many functions are increasingly the interest of study.

Yet, modern society has a complicated relationship with sleep health. The drive for productivity has led to spoken and unspoken pressures for minimizing sleep. Incentives can be both direct (e.g., rewards for staying awake to perform tasks) or indirect (e.g., being perceived as more industrious). Existing evidence suggests that Americans trade sleep hours for work or leisure hours (*10*) and that this trade-off is

[1]Department of Psychiatry, University of Arizona College of Medicine – Tucson, 1501 N. Campbell Ave., Suite AHSC 7326, Tucson, AZ 85724, USA. [2]Department of Psychology, University of Arizona College of Science, 1503 E. University Blvd., Room 507, Tucson, AZ 85721, USA.
*Corresponding author. Email: grandner@email.arizona.edu

exacerbated by modern sedentary lifestyles and use of electronic media at night. This larger discrepancy between the physiologic importance of sleep and its societal devaluing has deleterious consequences for health and well-being at the level of the individual, with collective effects that ripple across the population.

The field of neuroscience is well positioned to improve our basic understanding of how sleep is dynamically regulated and how it relates to other functions in the brain and periphery. Neuroscience is also well suited to leverage the translational potential of sleep to improve society. The fact that sleep is a fixture within so many aspects of our biology presents challenges and complications for the field. However, at the same time, it offers a range of opportunities for improving clinical outcomes in areas of medicine that are not accustomed to considering the influence of sleep on disease processes or the influence of social-environmental factors in shaping sleep health.

Conceptualizing sleep health

Sleep health, like proper nutrition, is not a unitary construct. Borrowing from paradigms historically used to define general health, Buysse's RU-SATED model outlines six dimensions of sleep health (*11*). These include:

1) Regularity. Regularity refers to the degree to which sleep occurs at approximately the same time each day. Irregular sleep schedules are an independent predictor of adverse health outcomes (*12*). These associations are thought to arise from decreased circadian synchronization across organ and tissue systems, disorganized patterns of cellular and organismal stress, and weekday-to-weekend fluctuations in sleep timing that may induce repetitive episodes of circadian disruption and occasional sleep insufficiency.

2) Satisfaction. Subjective sleep satisfaction remains an important indicator of sleep health. Although considered an imprecise measure, it predicts clinical outcomes (*13*), may reflect the presence of sleep impairments that are otherwise difficult to detect and assess, and may represent underlying physiologic processes for which there are no validated assessment tools, such as subcortical activation and local wakefulness in the brain (*14*).

3) Alertness. The ability to maintain attentive wakefulness during the day carries important safety implications, as well as implications for cardiometabolic health (*15*). Laboratory studies indicate that dulled alertness that stems from sleep loss can be profound and cumulative (*16*).

4) Timing. The circadian system partitions many aspects of sleep to the "biological night," which represents the portion of the 24-hour day that is behaviorally and physiologically conducive for sleep and sleep-related restorative activities organized within and between the brain and body (*17*).

5) Efficiency. Sleep efficiency refers to the ratio between the time spent asleep and the time available for sleeping. It is a standard metric used in the assessment and treatment of insomnia (*18*). Low sleep efficiency is associated with daytime dysfunction and adverse longitudinal health outcomes (*19*).

6) Duration. Consensus statements from several scientific, medical, and public health organizations recommend that adults, in general, should obtain at least 7 hours of sleep per night (*20*). Some statements have further advised that adults should not aim for more than 9 hours (*20*). Recommendations for children and adolescents have also been published (*20*, *21*), providing suggested ranges of daily sleep that scale inversely with age (i.e., infants require the most sleep, toddlers less sleep than infants, and so on until late adolescence when adultlike levels of sleep are recommended). For more than 50 years, self-reported sleep duration has been associated with mortality, including 7 of the 15 leading causes of death in the US (*22*).

Sleep and health outcomes

Cardiovascular health

Habitual sleep duration outside of the normative 7- to 8-hour range contributes to an increased incidence of hypertension, cardiovascular disease, and cardiovascular mortality (*23*, *24*). Cross-sectionally, habitual sleep duration and quality are associated with hypertension, hyperlipidemia, coronary disease, vascular disease, myocardial infarction, and stroke (*23*, *25*). Primary sleep disorders such as sleep apnea (*26*), insomnia (*24*, *27*), sleep-related movement disorders (*28*), and narcolepsy (*29*) also influence cardiovascular health. The mechanisms that bridge sleep to cardiovascular health remain elusive but seem to implicate sympathetic activation, cellular stress, inflammation, and metabolic dysregulation (*23*).

Metabolic health

Extensive investigation has revealed a cross-sectional relationship between habitual sleep duration outside the normative range and obesity-related outcomes (*30*). Meta-analyses of longitudinal studies suggest that the incidence of obesity is elevated in habitual short sleepers (≤6 hours) relative to 7- to 8-hour sleepers (*31*). This is consistent with other work showing that sleep loss impairs insulin and glucose homeostasis, disrupts metabolic hormones (e.g., leptin and ghrelin), alters adipose tissue function, and contributes to the incidence of type 2 diabetes mellitus (*32*).

Immunologic health

Sleep deprivation generates a proinflammatory state, as ascertained by tests performed in the lab or in the field (*33*). In laboratory settings, acute deprivation alters the expression, production, and/or function of cytokines, antibodies, and leukocytes. Field-based investigations suggest that overly short or long sleep duration promotes plasma-level markers of inflammation (*34*). Accordingly, habitual insufficient sleep has been linked to increased rates of infection, reduced recovery capacity, and blunted vaccine responses (*35*). The presence of primary sleep disorders also factors into immune health: Sleep apnea and insomnia are each associated with constitutive elevations in circulating immune markers (*36*, *37*). The relationship between sleep and the immune system is likely bidirectional (via homeostatic and neuroimmune interactions), with a deficiency in either poised to negatively affect the quality of the other (*38*).

Mental health

Insomnia is a well-established risk factor for depression and is implicated across several neuropsychiatric conditions, including anxiety, psychotic, and attention-deficit/hyperactivity disorders (*39*). Elevated suicide risk is associated with increased insomnia, nightmares, and habitual sleep duration outside of the normative range (*40*). Nocturnal wakefulness—simply being awake in the middle of the night—may also represent a risk factor for suicide (*41*, *42*). In general, sleep disruption produces basic functional impairments in emotion regulation, recognition, reasoning, and memory (*43*). This may be driven by neurophysiologic changes to cortical and subcortical structures that are unduly affected by acute or chronic sleep loss.

Behavioral health

Several studies have correlated differences in sleep to differences in the willingness of individuals to engage in unhealthy behaviors (especially those with an impulse-control component). For example, insufficient sleep and poor sleep quality have been associated with suboptimal dietary health at the population level (*44*). In the laboratory, acute sleep loss provokes increased calorie consumption, particularly late at night (*45*). Translational work in human and animal models suggests that eating during the night impairs metabolism (*46*). Insufficient sleep and poor sleep quality are further associated with more sedentary behavior, less exercise, more habitual smoking, and greater use of alcohol (*47*). Adding sleep health components to intervention strategies may aid in efforts such as smoking cessation (*48*).

Cognitive health

Connections between sleep and neurocognitive health are documented across the scientific literature. In addition to its acute effects on executive function and declarative memory (*49*), poor sleep has been prospectively linked with the development and progression of Alzheimer's

disease (AD) (*50*). A plausible mechanism linking sleep to AD is the function of the glymphatic system, which operates in a sleep-dependent fashion to clear AD-related proteins such as amyloid-β (Aβ) and tau from the brain's interstitial space. Sleep contributes to the regulation of Aβ and tau proteins and discourages their buildup into pathological aggregates (*51*). In the wider population, those who report habitual insufficient and/or poor sleep quality evince less work productivity (*52*), more drowsy driving (*53*), and increased propensity for injuries or errors (*54*).

Intersecting and interacting influences

Literature that links sleep to health is frequently circumscribed in that individual studies typically focus on a singular aspect of sleep health paired to one domain of interest. For example, studies of sleep duration and metabolic health do not typically quantify other sleep metrics (such as sleep quality, satisfaction, or daytime alertness) or assess multiple outcome domains (such as emotion regulation and decision-making). Yet, interactions between sleep health variables have been repeatedly demonstrated. For example, many individuals who experience insufficient sleep duration also experience insomnia symptoms (*55*). Understanding the isolated and combined effects of multiple domains of sleep health can potentially enhance our mechanistic understanding of clinical outcomes in depression, obesity, and those pertaining to cardiovascular events such as heart attack or stroke. Furthermore, although cardiometabolic and psychological domains are rarely assessed together, poor mental health can adversely affect cardiovascular, metabolic, and immune system activity, which are also influenced by neurocognitive processes such as health-related decision-making. Part of the mechanistic pathway that links poor sleep to cardiometabolic health may indeed be associated with mental health [e.g., emotional eating (*56*)] and/or altered executive function [e.g., decision-making about food choices (*57*)].

Conceptualizing sleep health in context

As outlined above, sleep health makes important contributions to a wide array of clinical outcomes. Additional work has conceptualized determinants of sleep itself. Understanding the factors that shape sleep health has the potential to improve intervention strategies and identify therapeutic targets applicable to cardiovascular, neurocognitive or neurodegenerative, and immune-related disorders such as cancer.

Individual-level factors

The most proximal influences on sleep are individual-level factors. These are factors that are embedded within the everyday life of the person—their demographic and personality traits, along with their beliefs, attitudes, and habits.

Age. Older adults experience systematic declines in sleep efficiency and time spent in slow-wave sleep (*58*). Age-related increases are observed in the prevalence of primary sleep disorders such as sleep apnea (*59*), chronic insomnia (*60*), nocturia (*61*), and movement-related nocturnal disorders [e.g., restless legs syndrome (*62*)]. Paradoxically, subjective measures of general sleep disturbance (*63*), perceived sleep insufficiency (*64*), and sleep debt (*65*) all tend to decrease. In a similar vein, short sleep duration is less strongly associated—or not associated at all—with obesity and hypertension in later life (*66*, *67*). Laboratory studies indicate that the cognitive performance of older individuals may also be more resilient to the detrimental effects of sleep loss (*68*).

Sex at birth. Females are at greater risk for insomnia and will report more dissatisfaction with sleep (*69*), whereas males are at relatively greater risk of sleep apnea (*59*). Specific to females are prolonged episodes of sleep disturbance and daytime fatigue associated with menstruation, pregnancy, the postpartum period, perimenopause, and menopause (*70*, *71*).

Race and ethnicity. Sleep disparities are well characterized in the US population (*72*–*75*). Black and African American individuals are significantly more likely to report insufficient sleep, as well as long sleep duration, relative to non-Hispanic white individuals. Various disparities related to sleep health are also seen in Hispanic and Latinx individuals, Asian and Asian American people, and multiracial individuals. These associations are reliable even after adjustment for socioeconomics and other demographic considerations (*76*). Notably, the relationship between sleep and cardiometabolic health that pervades the general population is differentiated by race and ethnicity, which suggests that disparities in sleep health may be integral to larger health care disparities (*77*). Future research priorities will aim to elucidate and address the racial and ethnic sleep disparities that figure more and more prominently in the literature (*78*).

Beliefs, attitudes, and habits. Work in this area is still emerging, though evidence suggests that individuals with more-positive beliefs and attitudes about sleep or who adopt sleep-protective behaviors (e.g., good sleep hygiene) are more likely to experience better sleep health (*79*).

Social-level factors

Individual-level factors are embedded within social-level factors. Examples of social-level factors that can potentially influence sleep are work and occupation, family and home, neighborhood, religion, social networks, and culture. Of these, several have been specifically evaluated with respect to sleep health.

Socioeconomic position. Associations abound between socioeconomic position and sleep health. Poverty (*75*, *80*) and the trappings of poverty, such as food insecurity (*76*, *81*), housing insecurity (*82*), and lower educational attainment (*72*, *80*), are all associated with worse measures of sleep, including insufficient sleep duration and decreased sleep quality.

Work and occupation. Shift work is a well-established impediment to healthy sleep (*83*). For non-shift workers, long commute times, stressful work conditions, and longer work hours impair sleep (*84*, *85*). Job loss and unemployment likewise moderate sleep health (*86*).

Home and family. Caregiving responsibilities present barriers to sleep health (*87*). Larger households may also contribute to patterns of insufficient sleep (*64*).

Neighborhood. Increased neighborhood light at night, increased traffic, and noise pollution are associated with worse sleep outcomes (*88*). People who live in neighborhoods that are more disordered (e.g., with more graffiti and crime) or perceived as less safe experience poorer sleep overall (e.g., shorter duration, less efficiency, and more daytime sleepiness) (*89*). Notably, individuals who perceive their neighborhood to be less socially cohesive also experience reduced sleep quality (*90*).

Societal-level factors

Societal-level factors are built from social-level factors. Societal factors include technology, public policy (*91*), geography (*92*), racism and discrimination (*93*), globalization and "24/7" society (i.e., general lifestyles of work and leisure in which people may be active at any hour and do not necessarily sleep consistently at night). Although these factors are not readily manipulable, recognizing their influence on sleep health can inform our understanding of how sleep is shaped and how this influence has the potential to shape other health- and disease-related outcomes.

Social-ecological model of sleep health

The embedded model of individual-, social-, and societal-level factors of sleep health, and the mental and physical health effects of sleep, are depicted in Fig. 1. This social-ecological model of sleep health was originally described in 2010 (*22*) and has since been conceptually refined (*94*). This model provides a framework for understanding sleep health in a broader biological and social-environmental context. The framework can be leveraged to inform better diagnosis, monitoring, and treatment of chronic diseases whose progression and treatment receptibility are often influenced by sleep.

Sleep health technology

One of the societal-level factors that has profoundly affected population sleep health is

Sleep health in context

There are several dimensions to sleep health, such as regularity, satisfaction, alertness, timing, efficiency, and duration. Alone and in combination, these dimensions interact with many aspects of our general health, including cardiovascular, metabolic, immune, mental, behavioral, and cognitive health. Interactions between sleep and health are further shaped by individual- and social-level factors operating within larger societal factors. All occur in the context of the day-night cycle.

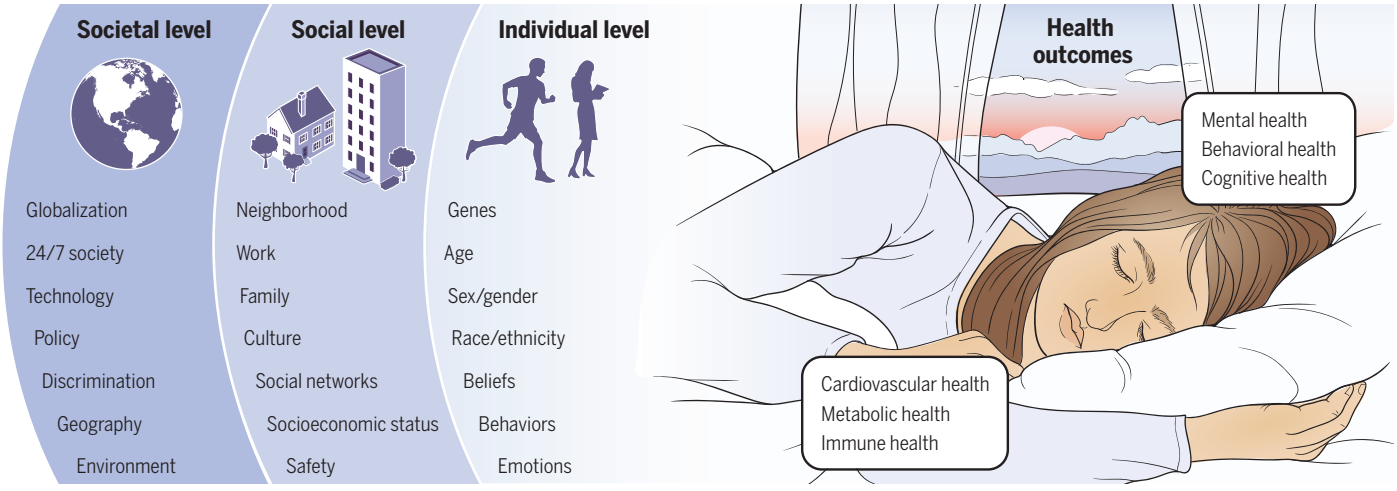


Fig. 1. Social-ecological model of sleep health. The right side of the image, which depicts a sleeping person, reinforces the importance of situational context in sleep physiology (*94*). The external environment is shown as a dawn-to-dusk transition, reinforcing the importance of day-night circadian context in sleep physiology (*6*). Health outcomes associated with sleep—cardiovascular health (*23*), metabolic health (*31*), immune health (*38*), mental health (*39*), behavioral health (*24*), and cognitive health (*50*)—are listed. On the left side of the image, the levels of the social-ecological model are depicted. Individual-level factors that influence sleep include age (*58*), genetics (*105*), sex and gender (*69*), beliefs (*94*), feelings and attitudes (*94*), race and ethnicity (*78*), and behaviors (*94*). These are embedded within social-level factors, which include neighborhood (*88*), work and occupation (*84*), family and home (*101*), culture (*80*), social networks (*90*), socioeconomics (*76*), and safety and security (*89*). These factors are further embedded within those at the societal level, such as such as technology (*95*), globalization (*94*), racism and discrimination (*93*), public policy (*91*), geography (*92*), the physical environment (*94*), and "24/7" society (*83*).

technological progress, particularly with regard to the proliferation of portable sleep-assessment technology, the ubiquitous access to digital media at night, and the cultivation of "big data" sources enabling access to sleep health information.

Portable sleep-assessment technology

Over the past few decades, wearable sleep actigraphy has undergone notable innovation, with newly built capabilities for multiaxial accelerometry and microelectromechanical systems enabling high-resolution detection of movement, as well as detection strategies for other peripheral signals (such as heart rate) that can improve assessments of sleep when computed alongside movement information (*95*). These devices are now available as consumer devices, some of which may be more accurate than their traditional scientific counterparts (*96*). The increasing use of these devices to assess sleep has led to clinical issues such as orthosomnia (*97*).

Electronic media use at night

Many people bring electronic devices into the bedroom and use them in bed during the nighttime hours that overlap with their routine sleep schedule. These trends are a cause for concern, because electronic media consumption at night prior to sleep can delay sleep onset, reduce sleep duration, impair sleep quality, worsen circadian health, and decrease daytime functioning (*98*). The mechanisms behind these disturbances range from exposure to short-wavelength light (*99*) to the cognitive and emotional stimulation caused by engaging content (*100*). These effects are pronounced in younger individuals, who have delayed sleep phases (*101*).

Big data and internet of things

As wearable sensor technology becomes increasingly inexpensive and sensors capable of inferring aspects of sleep-wake behavior become better integrated within the built environment, broader datasets will be collected to provide actionable knowledge about real-world sleep and circadian health. For example, large datasets curated from consumer apps have been used to describe changing sleep patterns during the COVID-19 pandemic (*102*); electronic medical records are providing insights at the provider, network, and national levels (*103*); and deep repositories of sleep-assessment data are being used to train artificial intelligence to better interpret sleep-wake physiology in relation to health (*104*).

Challenges and opportunities for translating sleep considerations into clinical study and practice

Challenges

1) Define the deeper biological context of sleep. Hallmarks for defining sleep include electrophysiological changes in the frequency, amplitude, and synchronicity of brain activity measured via electroencephalography (EEG) in humans and other mammals or local field potential recordings in invertebrates. Implicit within these definitions is that sleep is a phenomenon that can be objectively quantified only through networks of cells coordinated, in animals, with centralized nervous systems. Thus, these criteria cannot be applied universally across simpler organisms, such as those that populate the gut microbiome, or to single cells and cell masses involved with tumor growth and metastasis. There may be ways to define sleep at the single-cell level within the nucleus that have yet to be fully considered (*105*). The interphase nucleus of cells is a vibrant area, where chromosomes exhibit a polarized, multilayered spatial architecture on the basis of their proximity to structural elements of the nucleus, subnuclear structures, and one another along topologically associating domains and long-range chromatin interactions. These organization motifs—and many finer ones—contribute to how genomic expression is circadian-regulated and may provide insights into how sleep is expressed at the single-cell level or dysregulated in cancer cells or cells situated within various neurodegenerative disease contexts. These changes can become treatment targets in their own right.

ILLUSTRATION: KELLIE HOLOSKI/*SCIENCE*

2) Clarify the neuroscientific mechanisms that underlie different facets of sleep health. Although many studies have linked different aspects of sleep to outcomes, there is relatively little work that delineates the specific pathways by which this occurs. Perhaps answers lie in genetic studies. For example, people with an autosomal dominant mutation in the β1-adrenergic receptor gene (*ADRB1-A187V*) (*106*), a missense mutation in the *DEC2* gene (*BHLHE41*) (*107*), or a point mutation in neuropeptide S receptor 1 gene (*NPSR1*) (*108*) seem to require less sleep than those in the general population. Further characterization of these genes may help uncover presently unknown mechanistic pathways for sleep duration and increase the impetus to identify genes that influence other dimensions of sleep health, such as sleep efficiency and satisfaction. Improved understanding of the overarching "genetics of sleep health" would enable clinical investigators to develop genetic (along with social-environmental) profiles for the sleep health of trial participants and patients, as well as to consider when stronger or weaker associations might be expected between sleep and health outcomes of interest.

3) Identify translational models that link sleep and circadian neuroscience to clinical outcomes with wider population relevance. Efforts to study other generalizable contexts that connect sleep to health are needed. One example is in neurodevelopment: A mother's relative inability to sleep during pregnancy may be linked to increased adiposity in offspring (*109*), along with increased infant snoring (*110*), worse temperament (mood regulation) (*110*, *111*), and more nighttime awakenings before consolidation of the sleep schedule (*111*). Given these observations, work in animal models might explore the biological mechanisms by which sleep disruption during pregnancy can influence long-term risk prevalence for physical and mental health conditions. Perhaps maternal sleep deficiency leads to changes in the uterine environment, thus provoking epigenetic reprogramming of sleep and arousal in offspring (*112*). Consequences related to maternal sleep deprivation have already been reported in rodents (*113*). Greater study of prenatal sleep in parents and neonatal sleep in infants may produce insights into conditions with important neurodevelopmental components, such as intellectual disability, autism, and schizophrenia.

4) Integrate the behavioral, social, and environmental determinants of sleep health into neuroscientific models that leverage a social-ecological framework. A future challenge in the sleep field will be to incorporate what we are learning at the population level, by monitoring morbidity reports and the effects of public health policy, and then translate it back to preclinical animal models. Such an effort would require closer collaboration between human and animal researchers to create reliable paradigms for translating findings between species with overlapping, but in many cases separate, determinants of sleep health. This issue is as difficult as it is urgent. Attempts at addressing it might start with the simple premise that rodent sleep in the lab is often socially impoverished relative to sleep in the wild or the sleep of humans. Cosleeping might be a factor that better unmasks relationships between sleep and health in lab animals, thus enabling more reliable application of findings to people.

5) Develop better technological strategies for tracking sleep health and translating these data into actionable insights (with all due legal and ethical considerations). Many of the challenges associated with translating real-world sleep into personalized health interventions involve engineering and computational challenges, such as developing better sensors, developing better strategies for combining and integrating multimodal time-series data, and applying artificial intelligence to predict changes in sleep metrics that will have tangible effects on cardiovascular, immune, and brain function. Additional challenges include reconceptualizing the role of measurement in interventions, as well as reconceptualizing the measurements themselves (*95*). For example, a better understanding of existing signals (such as EEG or photoplethysmography) could lead to additional physiological insights about sleep under real-world conditions and—within the home environment—could delineate how sleep trajectories change upon disease onset and progression, as well as during intervention attempts timed to different disease stages.

Opportunities

1) Sleep plays a fundamental role in human physiology. Knowledge about sleep can potentially improve pain management, chronic disease treatment, and cognitive outcomes in neurodegenerative disorders, along with coordinating drug bioavailability. For these indications and others, a new technology model can be envisioned for the bedroom as a treatment site within the home that becomes operational as we fall asleep. Applied sleep neuroscience should explore how sleep-wake rhythms can empower individuals to better respond to clinical interventions, as well as help individuals recover from each day's pressures, thereby preparing them for tomorrow's disease risk.

2) Sleep is associated with many aspects of mental and physical well-being. Although sleep is acknowledged as a major contributor to mental and physical health outcomes, this growing understanding has yet to be codified into NIH and other federal guidelines that would encourage clinical trial designers to incorporate sleep as a common variable of interest and to formalize the scheduling, collection, and reporting of biosample collection with respect to time of day and an individual's sleep phase. With such mandates in place along with open science practices, biomedical research could synthesize information from multiple levels of analysis to provide insights into the fundamental contributions of sleep to individual organ systems and emergent physiology, as well as how sleep modifies the course of disease and response to experimental treatments.

3) Sleep health itself is multidimensional. Many investigations concerning human sleep are limited by an overly narrow focus on sleep deprivation and primary sleep disorders, thus creating a false dichotomy between typical and atypical sleep. The very concept of sleep health, which moves beyond clinical disorders to emphasize the positive contributions of sleep to mental and physical well-being, suggests that all sleep gradations are relevant to health outcomes. To the extent that this is the case, each aspect of sleep can become a "lever arm" in neuroscience to perpetuate good health or improve disease-related outcomes.

4) Sleep health may represent a pathway for reducing health disparities. Structural barriers, working independently or collectively, are known to have adverse effects on sleep, with likely effects on mental and physical health outcomes that enhance racial and ethnic disparities in health care. Reductions in health disparities may be aided with community-level engagement, investments in social services, and data-driven policies that (i) increase awareness of the importance of sleep, (ii) create greater vigilance for primary sleep disorders, and (iii) encourage discussion of sleep problems with medical professionals. Public campaigns such as these offer the opportunity for sleep neuroscience to make a tangible difference in the day-to-day lives of people in greatest need of help.

5) Sleep is becoming easier and less expensive to assess in the real world. The evolution of sleep measurement outside the confines of the laboratory presents many opportunities to harness an individual's own data in the service of personalized medical approaches that can improve the ease with which a person's sleep is conceptualized in relation to their health. Artificial intelligence and other big-data analytics can also examine sleep at the cohort or population scale, thus improving the availability and utility of data gathered across multiple naturalistic contexts (*104*).

Conclusion

Future neuroscience research on sleep health presents challenges and opportunities that will be difficult to untangle. For instance, many social-environmental determinants of

sleep health are not presently considered by medical professionals when placing an individual's sleep in context in the clinic, thus creating ambiguity as to whether sleep differences are the product of disease or reflect the aggregated circumstances of the individual (e.g., age, race, gender, socioeconomic status, work and home environment). Moreover, although sleep is acknowledged in the scientific literature as making decisive contributions to normal and abnormal physiology, it has yet to be standardized in clinical trial design. Perhaps nowhere is this more salient than in AD trials, which exemplify lost opportunities to benchmark the effects of amyloid-lowering drugs with respect to a person's previous and recent sleep history. Similar lost opportunities may be realized (in hindsight) for other medical conditions. In the present Review, we have attempted to itemize the core aspects of sleep health, describe the social-environmental determinants that contribute to each, and summarize how sleep ultimately factors into cardiovascular, metabolic, immunologic, mental, cognitive, and behavioral health. Along the way, we have pointed out areas of inquiry in neuroscience with the potential to accelerate the translation of sleep research from the lab to clinical practice. From distinguishing sleep states in single cells to understanding the genetics of sleep health, as well as from creating coherent translational bridges between lab animal and human sleep to creating credible and accessible sleep tracking technology in the home environment, neuroscience will be pivotal for integrating sleep health considerations into how we diagnose, monitor, treat, and test interventions for human brain disorders and other diseases. With any luck, the "chimera" of sleep can ultimately be leveraged at multiple interconnected levels to improve the mind and body of the individual, the collective well-being of society, and the essentially valuable parts of human existence.

REFERENCES AND NOTES

1. W. J. Joiner, *Physiology* **33**, 317–327 (2018).
2. C. B. Saper, P. M. Fuller, N. P. Pedersen, J. Lu, T. E. Scammell, *Neuron* **68**, 1023–1042 (2010).
3. P. Franken, *Curr. Opin. Neurobiol.* **23**, 864–872 (2013).
4. B. Rasch, J. Born, *Physiol. Rev.* **93**, 681–766 (2013).
5. G. Yankova, O. Bogomyakova, A. Tulupov, *Rev. Neurosci.* 10.1515/revneuro-2020-0106 (2021).
6. R. Allada, J. Bass, *N. Engl. J. Med.* **384**, 550–561 (2021).
7. A. Rechtschaffen, B. M. Bergmann, *Behav. Brain Res.* **69**, 55–63 (1995).
8. D. F. Dinges *et al*., *Sleep* **20**, 267–277 (1997).
9. B. C. Tefft, *Accid. Anal. Prev.* **45**, 180–186 (2012).
10. M. Basner *et al*., *Sleep* **30**, 1085–1095 (2007).
11. D. J. Buysse, *Sleep* **37**, 9–17 (2014).
12. T. Nakanishi-Minami, K. Kishida, T. Funahashi, I. Shimomura, *Diabetol. Metab. Syndr.* **4**, 18 (2012).
13. J. D. Edinger *et al*., *J. Clin. Sleep Med.* **11**, 311–334 (2015).
14. M. A. Grandner, M. L. Perlis, *JAMA Netw. Open* **2**, e1918214 (2019).
15. T. J. Balkin, T. Rupp, D. Picchioni, N. J. Wesensten, *Chest* **134**, 653–660 (2008).
16. S. Banks, H. P. Van Dongen, G. Maislin, D. F. Dinges, *Sleep* **33**, 1013–1026 (2010).
17. R. G. Foster, *Interface Focus* **10**, 20190098 (2020).
18. S. Schutte-Rodin, L. Broch, D. Buysse, C. Dorsey, M. Sateia, *J. Clin. Sleep Med.* **4**, 487–504 (2008).
19. E. M. Cespedes Feliciano *et al*., *Pediatrics* **142**, e20174085 (2018).
20. M. Hirshkowitz *et al*., *Sleep Health* **1**, 233–243 (2015).
21. S. Paruthi *et al*., *J. Clin. Sleep Med.* **12**, 785–786 (2016).
22. M. A. Grandner, L. Hale, M. Moore, N. P. Patel, *Sleep Med. Rev.* **14**, 191–203 (2010).
23. M. A. Grandner *et al*., *Curr. Opin. Cardiol.* **31**, 551–565 (2016).
24. L. Meng, Y. Zheng, R. Hui, *Hypertens. Res.* **36**, 985–995 (2013).
25. M. P. St-Onge *et al*., *Circulation* **134**, e367–e386 (2016).
26. J. M. Bock, S. Vungarala, S. Karim, V. K. Somers, *Can. J. Cardiol.* **37**, 756–765 (2021).
27. F. Sofi *et al*., *Eur. J. Prev. Cardiol.* **21**, 57–64 (2014).
28. I. C. Hwang, K. S. Na, Y. J. Lee, S. G. Kang, *Psychiatry Investig.* **15**, 701–709 (2018).
29. P. J. Jennum, G. Plazzi, A. Silvani, L. A. Surkin, Y. Dauvilliers, *Sleep Med. Rev.* **58**, 101440 (2021).
30. M. A. Grandner, *Sleep Health* **3**, 393–400 (2017).
31. Y. Wu, L. Zhai, D. Zhang, *Sleep Med.* **15**, 1456–1462 (2014).
32. D. Koren, E. M. Taveras, *Metabolism* **84**, 67–75 (2018).
33. L. Besedovsky, T. Lange, J. Born, *Pflugers Arch.* **463**, 121–137 (2012).
34. M. A. Grandner, M. R. Sands-Lincoln, V. M. Pak, S. N. Garland, *Nat. Sci. Sleep* **5**, 93–107 (2013).
35. A. A. Prather, S. D. Pressman, G. E. Miller, S. Cohen, *Int. J. Behav. Med.* **28**, 151–158 (2021).
36. V. M. Pak, M. A. Grandner, A. I. Pack, *Sleep Med. Rev.* **18**, 25–34 (2014).
37. M. R. Irwin, R. Olmstead, J. E. Carroll, *Biol. Psychiatry* **80**, 40–52 (2016).
38. L. Besedovsky, T. Lange, M. Haack, *Physiol. Rev.* **99**, 1325–1380 (2019).
39. E. Hertenstein *et al*., *Sleep Med. Rev.* **43**, 96–105 (2019).
40. M. L. Perlis *et al*., *Sleep Med. Rev.* **29**, 101–107 (2016).
41. A. S. Tubbs *et al*., *J. Clin. Psychiatry* **81**, 19m12964 (2020).
42. A. S. Tubbs, F. X. Fernandez, D. A. Johnson, M. L. Perlis, M. A. Grandner, *J. Clin. Psychiatry* **82**, 20m13820 (2021).
43. A. N. Goldstein, M. P. Walker, *Annu. Rev. Clin. Psychol.* **10**, 679–708 (2014).
44. X. Ji, M. A. Grandner, J. Liu, *Public Health Nutr.* **20**, 687–701 (2017).
45. R. R. Markwald *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5695–5700 (2013).
46. A. Chaix, E. N. C. Manoogian, G. C. Melkani, S. Panda, *Annu. Rev. Nutr.* **39**, 291–315 (2019).
47. F. Patterson, S. K. Malone, A. Lozano, M. A. Grandner, A. L. Hanlon, *Ann. Behav. Med.* **50**, 715–726 (2016).
48. F. Patterson *et al*., *Nicotine Tob. Res.* **21**, 139–148 (2019).
49. W. D. S. Killgore, in *Human Sleep and Cognition, Part I: Basic Research*, G. A. Kerkhof, H. P. A. Van Dongen, Eds. (Elsevier, 2010), chap. 7, pp. 105–129.
50. A. P. Spira, L. P. Chen-Edinboro, M. N. Wu, K. Yaffe, *Curr. Opin. Psychiatry* **27**, 478–483 (2014).
51. C. Wang, D. M. Holtzman, *Neuropsychopharmacology* **45**, 104–120 (2020).
52. S. K. Hui, M. A. Grandner, *J. Occup. Environ. Med.* **57**, 1031–1038 (2015).
53. Q. Maia, M. A. Grandner, J. Findley, I. Gurubhagavatula, *Accid. Anal. Prev.* **59**, 618–622 (2013).
54. K. Uehli *et al*., *Sleep Med. Rev.* **18**, 61–73 (2014).
55. M. A. Grandner, D. F. Kripke, *Psychosom. Med.* **66**, 239–241 (2004).
56. S. B. Ghani *et al*., *Behav. Sleep Med.* 10.1080/15402002.2021.1902814 (2021).
57. S. M. Greer, A. N. Goldstein, M. P. Walker, *Nat. Commun.* **4**, 2259 (2013).
58. M. M. Ohayon, M. A. Carskadon, C. Guilleminault, M. V. Vitiello, *Sleep* **27**, 1255–1273 (2004).
59. P. E. Peppard *et al*., *Am. J. Epidemiol.* **177**, 1006–1014 (2013).
60. S. Ancoli-Israel, *Sleep Med.* **10** (suppl. 1), S7–S11 (2009).
61. R. Asplund, in *Principles and Practice of Geriatric Sleep Medicine*, S. R. Pandi-Perumal, J. R. Monti, A. A. Monjan, Eds. (Cambridge Univ. Press, 2010), chap. 14, pp. 150–159.
62. P. Yeh, A. S. Walters, J. W. Tsuang, *Sleep Breath.* **16**, 987–1007 (2012).
63. I. Zilli, G. Ficca, P. Salzarulo, *Sleep Med.* **10**, 233–239 (2009).
64. M. A. Grandner *et al*., *Front. Neurol.* **6**, 112 (2015).
65. E. C. Fox *et al*., *Sleep Health* **4**, 317–324 (2018).
66. M. A. Grandner, E. A. Schopfer, M. Sands-Lincoln, N. Jackson, A. Malhotra, *Obesity* **23**, 2491–2498 (2015).
67. M. Grandner *et al*., *J. Clin. Sleep Med.* **14**, 1031–1039 (2018).
68. P. Philip *et al*., *J. Sleep Res.* **13**, 105–110 (2004).
69. M. H. Hall, C. E. Kline, S. Nowakowski, *F1000Prime Rep.* **7**, 63 (2015).
70. J. L. Shaver, N. F. Woods, *Menopause* **22**, 899–915 (2015).
71. L. Sweet, S. Arjyal, J. A. Kuller, S. Dotters-Katz, *Obstet. Gynecol. Surv.* **75**, 253–262 (2020).
72. F. A. Etindele Sosso, S. D. Holmes, A. A. Weinstein, *Sleep Health* **7**, 417–428 (2021).
73. S. Ahn *et al*., *Sleep Med.* **81**, 169–179 (2021).
74. L. Hale, D. P. Do, *Sleep* **30**, 1096–1103 (2007).
75. M. A. Grandner *et al*., *Sleep Med.* **11**, 470–478 (2010).
76. J. Whinnery, N. Jackson, P. Rattanaumpawan, M. A. Grandner, *Sleep* **37**, 601–611 (2014).
77. M. A. Grandner *et al*., *Sleep* **36**, 769–779 (2013).
78. C. L. Jackson, J. R. Walker, M. K. Brown, R. Das, N. L. Jones, *Sleep* **43**, zsaa037 (2020).
79. L. A. Irish, C. E. Kline, H. E. Gunn, D. J. Buysse, M. H. Hall, *Sleep Med. Rev.* **22**, 23–36 (2015).
80. M. A. Grandner, N. J. Williams, K. L. Knutson, D. Roberts, G. Jean-Louis, *Sleep Med.* **18**, 7–18 (2016).
81. S. M. Martinez, M. A. Grandner, A. Nazmi, E. R. Canedo, L. D. Ritchie, *Nutrients* **11**, 1419 (2019).
82. R. Bozick, W. M. Troxel, L. A. Karoly, *Sleep* **44**, zsab005 (2021).
83. S. Pallesen, B. Bjorvatn, S. Waage, A. Harris, D. Sagoe, *Front. Psychol.* **12**, 638252 (2021).
84. M. A. Grandner, *Am. J. Health Promot.* **32**, 1629–1634 (2018).
85. C. M. Barnes, N. F. Watson, *Sleep Med. Rev.* **47**, 112–118 (2019).
86. Q. D. Mai, T. D. Hill, L. Vila-Henninger, M. A. Grandner, *J. Sleep Res.* **28**, e12763 (2019).
87. E. Perez, P. B. Perrin, S. K. Lageman, T. Villaseñor, J. M. Dzierzewski, *Disabil. Rehabil.* 10.1080/09638288.2020.1814878 (2020).
88. J. C. Hunter, K. M. Hayden, *Public Health* **162**, 126–134 (2018).
89. T. E. Fuller-Rowell *et al*., *Sleep Med.* **81**, 341–349 (2021).
90. R. Robbins *et al*., *Sleep Med.* **60**, 165–172 (2019).
91. C. M. Barnes, C. L. Drake, *Perspect. Psychol. Sci.* **10**, 733–737 (2015).
92. M. A. Grandner *et al*., *Sleep Health* **1**, 158–165 (2015).
93. M. A. Grandner *et al*., *Behav. Sleep Med.* **10**, 235–249 (2012).
94. M. A. Grandner, *Sleep Med. Clin.* **15**, 319–340 (2020).
95. M. A. Grandner, M. R. Lujan, S. B. Ghani, *Sleep* **44**, zsab071 (2021).
96. E. D. Chinoy *et al*., *Sleep* **44**, zsaa291 (2021).
97. K. G. Baron, S. Abbott, N. Jao, N. Manalo, R. Mullen, *J. Clin. Sleep Med.* **13**, 351–354 (2017).
98. L. Hale *et al*., *Child Adolesc. Psychiatr. Clin. N. Am.* **27**, 229–245 (2018).
99. A. J. K. Phillips *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12019–12024 (2019).
100. E. Weaver, M. Gradisar, H. Dohnt, N. Lovato, P. Douglas, *J. Clin. Sleep Med.* **6**, 184–189 (2010).
101. M. A. Carskadon, *Pediatr. Clin. North Am.* **58**, 637–647 (2011).
102. N. Rezaei, M. A. Grandner, *Sleep Health* **7**, 303–313 (2021).
103. E. M. Wickwire, C. M. Morin, *Sleep Med. Rev.* **54**, 101387 (2020).
104. I. Perez-Pozuelo *et al*., *NPJ Digit. Med.* **3**, 42 (2020).
105. I. Pacheco-Bernal, F. Becerril-Pérez, L. Aguilar-Arnal, *Clin. Epigenetics* **11**, 79 (2019).
106. G. Shi *et al*., *Neuron* **103**, 1044–1055.e7 (2019).
107. A. Hirano *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3434–3439 (2018).
108. L. Xing *et al*., *Sci. Transl. Med.* **11**, eaax2014 (2019).
109. A. Brener *et al*., *Sci. Rep.* **10**, 13979 (2020).
110. R. Tauman *et al*., *Am. J. Obstet. Gynecol.* **212**, 656.e1–656.e7 (2015).
111. K. Nakahara *et al*., *Sci. Rep.* **10**, 11084 (2020).
112. L. Palagini, K. Biber, D. Riemann, *Sleep Med. Rev.* **18**, 225–235 (2014).
113. G. N. Pires *et al*., *J. Sleep Res.* **30**, e13135 (2021).

ACKNOWLEDGMENTS

The authors wish to thank L. Hale and C. Jackson for input on Fig. 1. **Funding:** This work was supported by grants from the National Institute on Drug Abuse (R01DA051321) and the National Institute on Minority Health and Health Disparities (R01MD011600). **Competing interests:** M.A.G. reports receiving grants from Jazz Pharmaceuticals, Kemin Foods, and CeraZ and performing consulting activities for Jazz Pharmaceuticals, Idorsia, Merck, Fitbit, Natrol, Smartypants Vitamins, Athleta, Nightfood, and Simple Habit in the past 3 years. F.-X.F. reports no potential conflicts of interest.

10.1126/science.abj8188

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



SEDIMENT TRANSPORT

Muddied waters

The climate of High Mountain Asia is becoming warmer and wetter. Li *et al.* present data showing that rivers originating in this region have experienced large increases in runoff and sediment fluxes over the past six decades, most dramatically since the mid-1990s. The authors project that sediment flux from those rivers could more than double by 2050 in the case of extreme climate change, with potentially serious impacts on the region's hydropower capacity, food security, and environment. —HJS

Science, abi9649, this issue p. 599

Rivers originating in High Mountain Asia, such as the Yangtze River, pictured here near Chongqing, China, have experienced increased sediment loading in recent decades.

LIPID TRANSPORT

Distinct mechanisms to block transporters

Transmembrane ATP-binding cassette (ABC) transporters are crucial cellular machines that move molecules small and large across membranes. In Gram-negative bacteria, outer membrane biogenesis is aided by an ABC transporter called MsbA, which flips lipopolysaccharide from the inner face of the cell membrane to the periplasmic face. Thélot *et al.* determined structures of two first-generation inhibitors bound to MsbA, TBT1 and G247, and found that they have distinct binding modes. Unlike most inhibitors, TBT1 triggers unproductive ATPase activity and induces a conformation similar to substrate bound. These structures will provide valuable information for the design of potential antimicrobial drugs. The authors have already identified a new lead compound from virtual screening based on the TBT1-induced conformation. —MAF

Science, abi9009, this issue p. 580

NUCLEAR ARCHITECTURE

Nuclear organization in the brain

The brain consists of different cell types, neurons and glial cells, that have different nuclear architecture. Takei *et al.* used multiplexed imaging tools to examine the spatial arrangement of more than 3000 DNA loci, along with epigenetic marks and gene expression patterns, simultaneously in the same single cells in mouse brain cortex. They observed features that are conserved across cell types, as well as cell-type-dependent spatial arrangements at the megabase level. At the level of tens of kilobases, they observed similar single-cell chromosome domain conformations in both active and inactive X chromosomes that are averaged out in population-based measurements. —DJ

Science, abj1966, this issue p. 586

MAGNETISM

Taking the measure of a magnet

The recent discovery of magnetism in two-dimensional (2D) materials has inspired efforts to understand its nature. Whereas the magnetism of monolayers of chromium iodide (CrI_3) can be understood in terms of out-of-plane magnetic anisotropy, the related material chromium chloride ($CrCl_3$) has spins that lie in the plane. Bedoya-Pinto *et al.* used molecular beam epitaxy to grow monolayers of $CrCl_3$ on graphene and studied its magnetic properties. Using x-ray magnetic circular dichroism measurements, the authors found that monolayer $CrCl_3$ is a ferromagnet, unlike bulk $CrCl_3$, which is antiferromagnetic. The scaling of the signal in the critical region indicated that the material belongs to the 2D-XY universality class, distinct from the Ising magnetism that some other 2D magnets exhibit. —JS

Science, abd5146, this issue p. 616

PHOTO: KEITMA /ALAMY STOCK PHOTO

PEROVSKITES

Stable emission in glass

Lead halide perovskites can exhibit bright, narrow band photoluminescence but have stability issues related to the formation of inactive phases and the loss of lead ions. Hou *et al.* show that the black, photoactive phase of cesium lead iodide can be stabilized by forming a composite with a glassy phase of a metal-organic framework through liquid-phase sintering. The photoluminescence is at least two orders of magnitude greater than that of the pure perovskite. The glass stabilizes the perovskite under high laser excitation, and about 80% of the photoluminescence was maintained after 10,000 hours of water immersion. —PDS

Science, abf4460, this issue p. 621

PHYTOPLANKTON

Circulating in the sunlit ocean

Marine plankton, which lie at the base of oceanic food chains, drive global biogeochemical fluxes, and knowledge of their distribution is key to understanding the response of oceans to environmental changes. Sommeria-Klein *et al.* explored the patterns and drivers of biogeography in eukaryotic plankton using a probabilistic model of taxon co-occurrence to compare the biogeography of 70 major groups, including a variety of size fractions and ecologies. The analysis is based on metabarcoding data from 129 stations in several oceanic provinces worldwide. Samples are from sunlit surface waters and, in about half of the stations, from the deep chlorophyll maximum. An essential message is that small phototrophs distribute mostly by latitude and bigger consumers are partitioned by ocean basin. —CA

Science, abb3717, this issue p. 594



Variations in the size of marine phytoplankton, here viewed under a microscope, influence their biogeographic distribution.

PHOTOS (LEFT TO RIGHT): DR. D.P. WILSON/SCIENCE SOURCE; ANT PHOTO LIBRARY/SCIENCE SOURCE

PLANT SCIENCE

Nodulation regulation

Legumes convert atmospheric nitrogen into biologically useful ammonium with the help of symbiotic bacteria housed in root nodules. Much of nodule development is controlled by the transcription factor NODULE INCEPTION (NIN). Feng *et al.* show that NIN is proteolytically processed to release a fragment that regulates the later stages of nodulation when the nodules acquire nitrogen-fixing capability. In related work, Jiang *et al.* identified members of the NIN-like protein (NLP) transcription factor family as being regulators of leghemoglobin expression acting through an unusual promotor motif shared across legumes. —PJH

Science, abg2804, abg5945, this issue p. 629, p. 625

NEUROSCIENCE

Flipping astrocytes to treat Alzheimer's

Astrocytes are implicated in β-amyloid accumulation in Alzheimer's disease. Raha *et al.* identified an orally ingested combination of clinically approved drugs that induced astrocytes to take up and break down β-amyloid fibrils instead of promoting β-amyloid buildup. Astrocytes in cell cultures and in Alzheimer's disease model mice were stimulated to phagocytose β-amyloid by retinoic acid and to degrade β-amyloid by the cholesterol-lowering drug gemfibrozil. In mice, the combination of retinoic acid and gemfibrozil reduced β-amyloid load and improved cognitive function. —LKF

Sci. Signal. **14**, eabg4747 (2021).

IN OTHER JOURNALS

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



PALEOECOLOGY

Prehistoric eggshells

Cassowaries are now restricted to the rainforests of Australia and New Guinea. They are among the few large ratite birds to have survived the global expansion of human populations, unlike the moas of New Zealand, which were hunted to extinction. Douglass *et al.* report archaeological evidence for human exploitation of cassowaries in the mountain forests of New Guinea in the Late Pleistocene and Early Holocene. Their data suggest that harvesting of cassowary eggs may have been preferred over the hunting of adult birds. Microstructural features of the sampled eggshells suggest that there was preferential exploitation of eggs close to hatching, which raises the intriguing possibility that humans may have attempted the rearing of cassowary chicks. –AMS

Proc. Natl Acad. Sci. U.S.A. **118**, e2100117118 (2021).

Archaeological evidence indicates that cassowary eggs were exploited by humans in New Guinea from the Late Pleistocene to the early Holocene.

MACHINE LEARNING

Bayesian optimization of electrocatalyst

In recent years, high-entropy alloys (HEAs) have attracted a lot of attention for catalytic applications. The vast composition space of HEAs ensures the existence of surface sites having optimal catalytic properties, yet such complexity represents a great challenge for combinatorial approaches to experimental catalyst discovery. Pedersen *et al.* constructed a methodology based on Bayesian optimization combined with kinetic modeling and density functional theory calculations to predict the most active compositions for the electrochemical oxygen reduction reaction from the quinary HEAs Ag-Ir-Pd-Pt-Ru and



BIOMINERALOGY

Precursor double down

Biominerals that make up shells, coral skeletons, and other structures often grow differently than other types of crystals. Instead of atom-by-atom growth, solid amorphous precursor particles attach and crystallize. However, Stifler *et al.* found evidence of liquid precursor particles from observations of sea urchin spines regenerating from cuts. Liquid-like pools were evident in these regenerating spines, along with other observations consistent with a liquid morphology. The authors tentatively conclude that these liquid droplets are precursors to the well-known solid amorphous calcium carbonate precursor particles that help to form the spines. —BG *Cryst. Growth Des.* 10.1021/acs.cgd.1c00865 (2021).

The solid amorphous calcium carbonate precursor particles from which the spines of the purple sea urchin (pictured) grow are themselves derived from liquid precursor particles.

Ir-Pd-Pt-Rh-Ru. The discovered optimal binary alloys have been successfully validated experimentally. The proposed methodology substantially reduces the total number of experiments required for optimizing the composition space, making catalyst design for HEAs practically feasible. —YS

Angew. Chem. Int. Ed. 10.1002/anie.202108116 (2021).

CELL BIOLOGY

Teaming up to finish a kill

Cytotoxic T lymphocytes (CTLs) are thought to kill target cells in a rapid, all-or-nothing process. This may be the case when target cell immunogenicity is high; however, in solid tumors, single CTL encounters may fail to kill target cells. Weigelin *et al.* found that migrating CTLs move between cancer cells and initiate apoptosis by additive sublethal interactions or "additive cytotoxicity." Sublethal damage occurs by perforin-dependent membrane pore formation, nuclear lamina rupture, and DNA disruption. In mouse melanoma tumors in vivo, additive cytotoxicity in interstitial tissue results in preferential elimination of invading tumor cells. –SMH

Nat. Commun. **12**, 5217 (2021).

SKIN INFLAMMATION

A new culprit in psoriasis

Chronic inflammatory skin diseases such as psoriasis and atopic dermatitis are caused by overactive immune responses. Nakamizo *et al.* discovered a previously unrecognized type of immune cell in the skin of individuals with psoriasis. By performing single-cell flow cytometry and RNA-sequencing analysis on specific immune cells called macrophages and dendritic cells (DCs), the researchers created a molecular signature for skin inflammation and report that $CD14^+$ type 3 DCs were more frequent in psoriasis lesions compared with healthy skin. This newly discovered DC subset expressed the cytokines *IL23A* and *IL1B*, which are known to drive inflammatory events in psoriasis. –PNK

J. Exp. Med. **218**, e20202345 (2021).

ATOM INTERFEROMETRY

Entanglement-enhanced atom interferometer

Atom interferometers work on the same principle as light interferometers but using the wave properties of atoms. Typically, a cloud of cold atoms is split, dropped into separate paths, and then brought back together to reveal interference fringes. Anders *et al.* demonstrate that a cloud of cold rubidium atoms cooled into a Bose-Einstein condensate, a quantum coherent state in which all of the atoms occupy the same quantum mechanical ground state, can enhance the sensitivity of the interferometer. Splitting the condensate into two momentum-entangled modes results in a sensitivity beyond the standard quantum limit. Using the setup in existing atom interferometers is expected to provide enhanced detection capability to test questions in fundamental physics, quantum gradiometers, and gravitational wave detectors. —ISO

Phys. Rev. Lett. **127**, 140402 (2021).

DEVELOPMENTAL BIOLOGY

Transitional metabolism

Various metabolites and metabolic pathways affect stem cell and embryo development. Zhao *et. al.* used a mass spectrometry–based method to measure metabolite abundance as mouse embryos transited between developmental stages. Combining metabolomics profiling with transcriptomics analysis revealed unique metabolic signatures as embryos developed from the totipotent to the pluripotent stage. Thus, as two-cell embryos become blastocysts, enzymes involved in the tricarboxylic acid cycle and oxidative phosphorylation increase activity and metabolite balance changes. For example, L-2-hydroxyglutarate in the two-cell embryos declines, resulting in global histone methylation remodeling during preimplantation development. –BAP

Nat. Metab. 10.1038/s42255-021-00464-x (2021).

PHOTO: RALPH PACE/MINDEN PICTURES

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

NEUROSCIENCE

The source of memory

Where are memories formed and stored? The answers to this question could have wide-ranging implications. In a Perspective, Shin *et al.* discuss their hypothesis that layer 1 of the neocortex is where long-term semantic memory is formed and stored. The pyramidal neurons that comprise the neocortex are ideally suited to this role because there are numerous inputs from other regions of the brain, allowing features and context to be programmed together in one cell. Understanding this process, as well as how memories are retrieved, could be informative for memory disorders, learning, and even how artificial intelligence networks are designed. —GKA

Science, abk1859, this issue p. 538

IMMUNOLOGY

The clock is ticking for senescent cells

Senescent cells promote their own recognition and removal through the immune system by generating a bioactive secretome called the senescence-associated secretory phenotype (SASP). Sturmlechner *et al*. report that the cell cycle regulator p21 directs an early form of the SASP, which they call the p21-activated secretory phenotype (PASP) (see the Perspective by Reen and Gil). As part of the PASP, the chemokine CXCL14 attracts macrophages, which monitor stressed cells expressing elevated p21. If stressed cells recuperate and p21 levels return to normal within 4 days, then macrophages disengage from their targets. Otherwise, macrophages recruit cytotoxic T cells that facilitate target cell removal. Other cell cycle regulators such as p16 can induce many factors overlapping with the PASP, but p21 uniquely drives this CXCL14-mediated "timer" mechanism of senescent cell immunosurveillance. —STS

Science, abb3420, this issue p. 577; see also abm3229, p. 534

INDIGENOUS PEOPLES

Long-term impacts of land dispossession

To date, we lack precise estimates of the extent to which Indigenous peoples in parts of North America were dispossessed of their lands and forced to migrate by colonial settlers, as well as how present-day tribal lands compare to their original lands. Farrell *et al*. constructed a new dataset within the boundaries of the current-day United States and found that Indigenous land density and spread in has been reduced by nearly 99% (see the Perspective by Fixico). Present-day tribal lands are more vulnerable to climate change and contain fewer resources. Research and policy implications of these findings are discussed. —TSR

Science, abe4943, this issue p. 578; see also abl6288, p. 536

CORONAVIRUS

The bat connection

The heterogeneity of COVID-19 makes it challenging to predict the course of infection in an individual. Upon virus infection, interferons (IFNs) generate the initial signals for cellular defenses. Knowing that defects in IFN signaling are associated with more severe COVID-19, Wickenhagen *et al*. used IFN-stimulated gene expression screening on human lung cells from which they identified a gene for 2′-5′-oligoadenylate synthetase 1 (OAS1) (see the Perspective by Schoggins). OAS1 stimulates RNase L to inhibit the virus with a surprising degree of specificity, targeting the membranous organelles in which it replicates. In most mammals, OAS1 is attached to membranes by a prenyl group. However, billions of humans do not have the prenylated OAS1 haplotype, including many experiencing severe COVID-19. The same is true for horseshoe bats, prolific sources of betacoronaviruses, because of an ancient retrotransposition event. —CA

Science, abj3624, this issue p. 579; see also abm3921, p. 535

POLYMER CHEMISTRY

Polyethylene with a nickel's worth of CO

The biggest problem with polyethylene, the most abundantly manufactured plastic, is that it doesn't break down easily once it is discarded. Chemists have long sought to introduce small quantities of carbon monoxide (CO) into polyethylene chains to promote photodegradation, but too much CO tends to jump in and spoil the plastic's other properties. Baur *et al*. report that nickel catalysts coordinated by bulky phosphinophenolate ligands can catalyze ethylene polymerization with approximately 1% CO incorporation, preserving tensile strength while promoting degradation under ultraviolet exposure (see the Perspective by Sobkowicz). —JSY

Science, abi8183, this issue p. 604; see also abm2306, p. 540

SUPERCONDUCTIVITY

Measuring the effective charge

At low enough temperatures, superconductors are capable of conducting electricity without any resistance because of the formation of so-called Cooper pairs of electrons. Cooper pairs typically form at the same critical temperature at which superconductivity sets in. In certain materials, they are thought to form above that temperature, but showing this property directly in an experiment is tricky. Bastiaans *et al*. used tunneling noise spectroscopy to measure the effective charge of current carriers in the disordered superconductor titanium nitride. As expected, below the critical temperature, the effective charge was equal to two electron charges. However, this behavior persisted above the critical temperature, indicating that electron pairs exist in that regime. —JS

Science, abe3987, this issue p. 608

PLASMONICS

Switching a polymer

Electrically switchable metasurfaces and plasmonic materials will enable the development of active nanophotonic technology. Karst *et al*. show that a metallic polymer can be used for electrical switching of plasmonic nanoantenna resonances. The plasmonic resonance can be completely switched ON and OFF with switching speeds up to 30 hertz (video rate), low switching voltages of ±1 volt (complementary metal-oxide semiconductor compatible), and a transmission modulation depth of 100%. The results could have applications in nanophotonic devices such as those used in augmented and virtual reality imaging applications. —ISO

Science, abj3433, this issue p. 612

SKIN INFLAMMATION

IL-12 curtails cutaneous inflammation

Psoriasis is primarily a type 3 immune response–driven pathology, motivating the development of interleukin-23 (IL-23)–neutralizing therapies. Despite the initial clinical success of monoclonal antibodies targeting the shared IL-12/IL-23 p40 subunit, the superior efficacy of IL-23–specific antibodies suggests that sparing IL-12 is beneficial. To investigate the cell-type-specific role of IL-12 signaling during psoriasiform inflammation, Zwicky

et al. developed mice with a conditional deletion of the IL-12 receptor beta 2-subunit gene (*Il12rb2*). Although *Il12rb2* was highly expressed in skin γδ T cells and a subset of interfollicular keratinocytes, mice with *Il12rb2* deleted in keratinocytes but not hematopoietic cells developed more severe skin inflammation in a model of psoriasis. IL-12 signaling restrained keratinocyte proliferation and counteracted IL-23/IL-17 cytokine signaling, highlighting the noninflammatory, regulatory function of IL-12 in skin pathologies dependent on type 3 immunity. —CO

Sci. Immunol. **6**, eabg9012 (2021).

TRANSPLANT MEDICINE

Inducing immune suppression

Type 1 regulatory T cells (Tr1 cells) are a population of inducible T cells that suppress graft-versus-host disease after allogeneic stem cell transplantation. To determine the mechanisms of suppression by Tr1 cells, Chen *et al*. compared Tr1 cells from a human Tr1-enriched cell product called T-allo10 with non-Tr1 cells from the same product. The authors found that Tr1 cells had a distinct T cell receptor repertoire and transcriptional program. Additionally, Tr1 cells suppressed through both interleukin-10 production and expression of two key coinhibitory proteins, CTLA-4 and PD-1. Finally, the authors showed in patients receiving T-allo10 that Tr1 cells persist up to 1 year after transfer. These findings further characterize the function of Tr1 cells and demonstrate their clinical relevance for graft-versus-host disease. —CSM

Sci. Transl. Med. **13**, eabf5264 (2021).

ARCHAEOLOGY

Identifying Tatanka Iyotake's descendents

The genetic analysis of the genealogies of historical figures is difficult because of the complexities of parental inheritance, restricted or limited access to samples, and the fact that the already limited amounts of DNA in such samples are often highly degraded. At the request of a presumed descendant of the famous Lakota leader Tatanka Iyotake (Sitting Bull), Moltke *et al.* developed a powerful new method for the analysis of DNA in these kinds of circumstances. Using a small piece of Sitting Bull's hair that had been repatriated to his presumed descendants in 2007, the team definitively identified one of them as Sitting Bull's great-grandson. The method shows promise for the analysis of samples with very small amounts of endogenous DNA. —MSA

Sci. Adv. 10.1126/sciadv.abh2013 (2021).

RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

p21 produces a bioactive secretome that places stressed cells under immunosurveillance

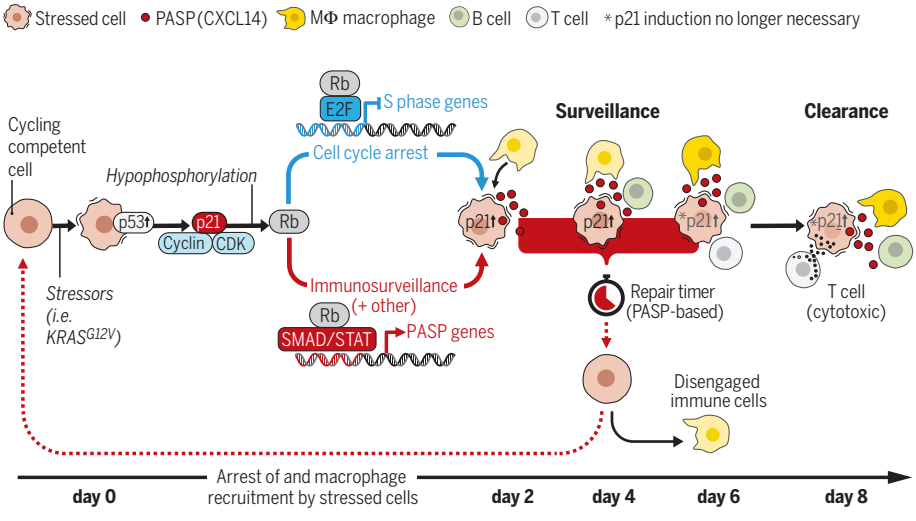
Ines Sturmlechner, Cheng Zhang, Chance C. Sine, Erik-Jan van Deursen, Karthik B. Jeganathan, Naomi Hamada, Jan Grasic, David Friedman, Jeremy T. Stutchman, Ismail Can, Masakazu Hamada, Do Young Lim, Jeong-Heon Lee, Tamas Ordog, Remi-Martin Laberge, Virginia Shapiro, Darren J. Baker, Hu Li*, Jan M. van Deursen*

INTRODUCTION: Complex multicellular organisms are subject to a myriad of cellular stresses, which can be managed through cell-intrinsic adaptation and repair mechanisms. Cells that fail to recover activate programs that lead to regulated cell death or cellular senescence, thereby limiting the risk of neoplastic transformation. Evidence is emerging that intercellular communication also plays an important role in dealing with cellular stresses. For instance, cells experiencing DNA damage display cell-surface ligands that facilitate lymphocyte recognition and secrete cytokines that attract myeloid cells. Furthermore, cells undergoing cellular senescence generate a bioactive secretome known as the senescence-associated secretory phenotype (SASP), which facilitates senescent cell (SNC) recognition by the immune system.

RATIONALE: Stressed cells that become senescent have been implicated in various biological processes beyond cancer, including development, tissue repair, aging, and age-related diseases, presumably through the paracrine actions of the SASP. To better understand the molecular properties of SNCs, we sought to identify key determinants of SNC identity by screening for genes nearby senescence-associated super-enhancers conserved across stressors, cell types, and mammalian species. Through this approach, we identified *p21* (*Cdkn1a*), which encodes the cyclin-dependent kinase (CDK) inhibitor p21, and conducted an in-depth analysis of its functions from the time cells first experience stress until they have become senescent.

RESULTS: We found that p21—in addition to its function in maintaining the cell-cycle arrest of SNCs—has a prominent role in establishing the SASP through retinoblastoma protein (Rb)–dependent transcription involving select SMAD and STAT transcription factors. Although this transcriptional program remains active in SNCs, p21 initiates this program as a first response to stress occurring in parallel with cell-cycle arrest. The resulting immediate-early secretome, which we term the p21-activated secretory phenotype (PASP), comprises several hundred factors, including the chemokine CXCL14, which recruits macrophages to surveil stressed cells with elevated p21. In mouse liver, these macrophages disengage if cells normalize p21 levels within 4 days after its induction. However, if p21 remains elevated, the adjoined macrophages polarize toward an M1 phenotype, and cytotoxic T lymphocytes arrive to execute target cell elimination when classical markers of senescence are not yet detectable. This scenario also occurs with oncogenic KRAS-mediated p21 induction, highlighting the physiological relevance of the uncovered immunosurveillance mechanism to tumor suppression. By contrast, CDK inhibitor p16, which is often elevated in SNCs and also halts cell-cycle progression through Rb hypophosphorylation, does not induce immunosurveillance when overexpressed in mouse liver. Although p16 induction yields an immediate-early secretome that consists largely of factors that overlap with the PASP, there are far fewer factors, and CXCL14 is absent. Studies on CDK inhibitor p27 further suggested that coordinated induction of cell-cycle arrest and immunosurveillance is a distinctive feature of p21.

CONCLUSION: Our study demonstrates that in response to cellular stress, p21 alters the transcription regulatory properties of Rb to not only inhibit genes required for cell-cycle progression but also activate a large collection of genes implicated in diverse biological functions, including immunosurveillance. By promptly recruiting macrophages to stressed cells, p21 sets a biological timer that allows for a period in which stressed cells can recuperate, thereby normalizing p21. This timer expires when the immune response transitions from a surveillance to a clearance mode. As such, p21 provides a first line of defense against dysfunctional cells that can become cancerous or otherwise cause pathology. Given that p21 is a p53 target gene, it will be important to define whether this p21-controlled immunoclearance mechanism is perturbed in cancer cells that lack functional p53, and if so, to explore the therapeutic impact of interventions that reactivate it. ■



p21 places stressed cells under immunosurveillance. p21 induced by cells under duress activates an Rb-dependent transcriptional program that produces CXCL14 and other secreted factors. CXCL14 promptly recruits macrophages, which disengage if stressed cells recuperate and normalize p21 within 4 days. However, if p21 induction persists, the adjoined macrophages polarize toward an M1 phenotype and cytotoxic T cells arrive to eliminate target cells, including preneoplastic cells.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: janvandeursen2@gmail.com (J.M.v.D.); li.hu@mayo.edu (H.L.)
Cite this article as I. Sturmlechner *et al.*, *Science* 374, eabb3420 (2021). DOI: 10.1126/science.abb3420

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abb3420

RESEARCH ARTICLE

IMMUNOLOGY

p21 produces a bioactive secretome that places stressed cells under immunosurveillance

Ines Sturmlechner[1,2], Cheng Zhang[3], Chance C. Sine[1], Erik-Jan van Deursen[1], Karthik B. Jeganathan[1], Naomi Hamada[1], Jan Grasic[1], David Friedman[4], Jeremy T. Stutchman[1], Ismail Can[5], Masakazu Hamada[1], Do Young Lim[1], Jeong-Heon Lee[6], Tamas Ordog[6,7], Remi-Martin Laberge[8], Virginia Shapiro[4], Darren J. Baker[1,5], Hu Li[3]*, Jan M. van Deursen[1,5]*

Immune cells identify and destroy damaged cells to prevent them from causing cancer or other pathologies by mechanisms that remain poorly understood. Here, we report that the cell-cycle inhibitor p21 places cells under immunosurveillance to establish a biological timer mechanism that controls cell fate. p21 activates retinoblastoma protein (Rb)–dependent transcription at select gene promoters to generate a complex bioactive secretome, termed p21-activated secretory phenotype (PASP). The PASP includes the chemokine CXCL14, which promptly attracts macrophages. These macrophages disengage if cells normalize p21 within 4 days, but if p21 induction persists, they polarize toward an M1 phenotype and lymphocytes mount a cytotoxic T cell response to eliminate target cells, including preneoplastic cells. Thus, p21 concurrently induces proliferative arrest and immunosurveillance of cells under duress.

Cells in complex multicellular organisms are subject to a myriad of stresses, which can be managed through various cell-autonomous adaptation and repair mechanisms (*1*). Cells that fail to recuperate—for instance, because of severe or prolonged stresses—engage programs that execute regulated cell death or cellular senescence, thereby limiting the risk of neoplastic transformation (*2*, *3*). The cellular senescence program is characterized by permanent cell-cycle withdrawal through the induction of cyclin-dependent kinase inhibitors like p21 and p16, encoded by *p21* (*CDKN1A*) and *p16* (*CDKN2A*), respectively (*4*–*7*). In mammals, cellular senescence has also been implicated in biological processes beyond cancer, including development (*8*, *9*), tissue repair (*10*), aging, and age-related disorders (*3*, *11*–*15*). Senescent cells presumably exert these functions by virtue of the senescence-associated secretory phenotype (SASP), a diverse collection of secreted factors (SFs) often enriched in immune-modulatory cytokines and chemokines, matrix remodeling enzymes, and growth factors (*5*, *16*–*18*). Once established, senescent cells can persist in tissues and organs for prolonged periods of time (*19*), although they can also be recognized and eliminated by lymphocytes through mechanisms that remain ill-defined (*4*, *16*, *20*).

We sought to better understand the properties of senescent cells at a molecular mechanistic level by identifying senescence-associated super-enhancer (SASE)–controlled genes. Super-enhancers are large enhancers that are highly enriched in certain chromatin modifications and bind large amounts of numerous transcription factors (TFs) and coactivators (*21*, *22*). Super-enhancers regulate genes with important functions in processes that are cell type–specific or define cell identity (*23*, *24*), which led us to hypothesize that the identification and in-depth characterization of genes that come under the control of super-enhancers with senescence would be particularly informative about the inner workings of this cell fate. In pursuing this idea, we focused our efforts on genes controlled by SASEs that are highly conserved across species, cell types, and senescence-inducing stressors. Our study illuminates that immediate-early cell-autonomous reactions to cellular stress are coordinated with cell nonautonomous responses through p21-mediated changes in the Rb-dependent transcriptional landscape to establish a biological timer mechanism for cell fate decisions.

p21 regulates the SASP through Rb-dependent transcription

In our initial screen, we exposed primary mouse embryonic fibroblasts (MEFs) to three distinct senescence-inducing stressors: γ-irradiation (IR), extensive replication (REP), and oncogene-induced (OI) signaling by overexpression of human $KRAS^{G12V}$ (fig. S1). We mapped the common super-enhancer changes as these cells transitioned to a senescent state, and identified the transcriptionally activated genes associated with these super-enhancers (fig. S2, A and B). We uncovered 50 such genes (fig. S2B and tables S1 and S2), three of which were also associated with a SASE in IR-senescent human fetal lung (IMR-90) cells and transcriptionally up-regulated in both IR- and OI-senescent IMR-90 cells, including *p21* (fig. S2, A to C, and table S3). Notably, H3K27Ac chromatin immunoprecipitation–quantitative polymerase chain reaction (ChIP-qPCR) on OI-senescent cells (SNCs) collected from mouse liver indicated that the SASE identified near the *p21* locus was conserved in vivo (fig. S2, D to J).

Fully SNCs deficient in p21 incorporated 5-ethynyl-2′-deoxyuridine (EdU) (fig. S3, A to D), indicating that sustaining p21 in the senescent state is important to prevent cell-cycle reentry through continued transcriptional repression of E2F target genes via hypophosphorylation of Rb (*25*). Intriguingly, knocking down *p21* in SNCs also decreased expression of multiple SASP factors as determined by reverse transcription (RT)–qPCR for a panel of well-established SASP factors (fig. S3E). Comprehensive transcriptomic analysis of IR-senescent MEFs using RNA sequencing (RNA-seq) revealed that about a third of the SASP (188 of 503 factors) was p21-dependent (Fig. 1A; fig. S4, A to C; and table S4). Similarly, nearly half the SASP (167 of 354 factors) identified in IR-senescent IMR-90 cells was dependent on p21 (Fig. 1A, fig. S4D, and table S4), which prompted us to probe the mechanism(s) underlying these p21-dependent secretory phenotypes, hereafter referred to as the p21-activated secretory phenotype (PASP).

We first focused on Rb and found that its absence in SNCs not only activated E2F target genes (figs. S3, F to I, and S5) but also decreased expression of most of the SASP factors down-regulated with *p21* deficiency (Fig. 1, A and B; fig. S3J; and table S4), suggesting that p21 confers its effect on the SASP through hypophosphorylation of Rb. To explore how p21-mediated Rb hypophosphorylation might activate SASP genes, we identified TFs that have been linked to the SASP, inflammation, or cytokine production and used their transcriptional targets in overrepresentation analyses on RNA-seq data from IR-, REP, OI-senescent MEFs, IR-senescent IMR-90 cells, and their non-senescent counterparts. We found that RELA, CEBPβ, SMAD2, SMAD3, STAT1, STAT5A/B, and STAT6 were consistently more transcriptionally active in SNCs than in non-SNCs regardless of senescence-inducing stressor or species (Fig. 1C). RELA, SMAD2, SMAD3, STAT1, and STAT6 lost this status when *p21* or *Rb* was depleted (Fig. 1C). Thus, hypophosphorylated

[1]Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN, USA. [2]Department of Pediatrics, Molecular Genetics Section, University of Groningen, University Medical Center Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, Netherlands. [3]Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA. [4]Department of Immunology, Mayo Clinic, Rochester, MN, USA. [5]Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, USA. [6]Epigenomics Program, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA. [7]Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN, USA. [8]Unity Biotechnology, South San Francisco, CA 94080, USA.
*Corresponding author. Email: janvandeursen2@gmail.com (J.M.v.D.); li.hu@mayo.edu (H.L.)

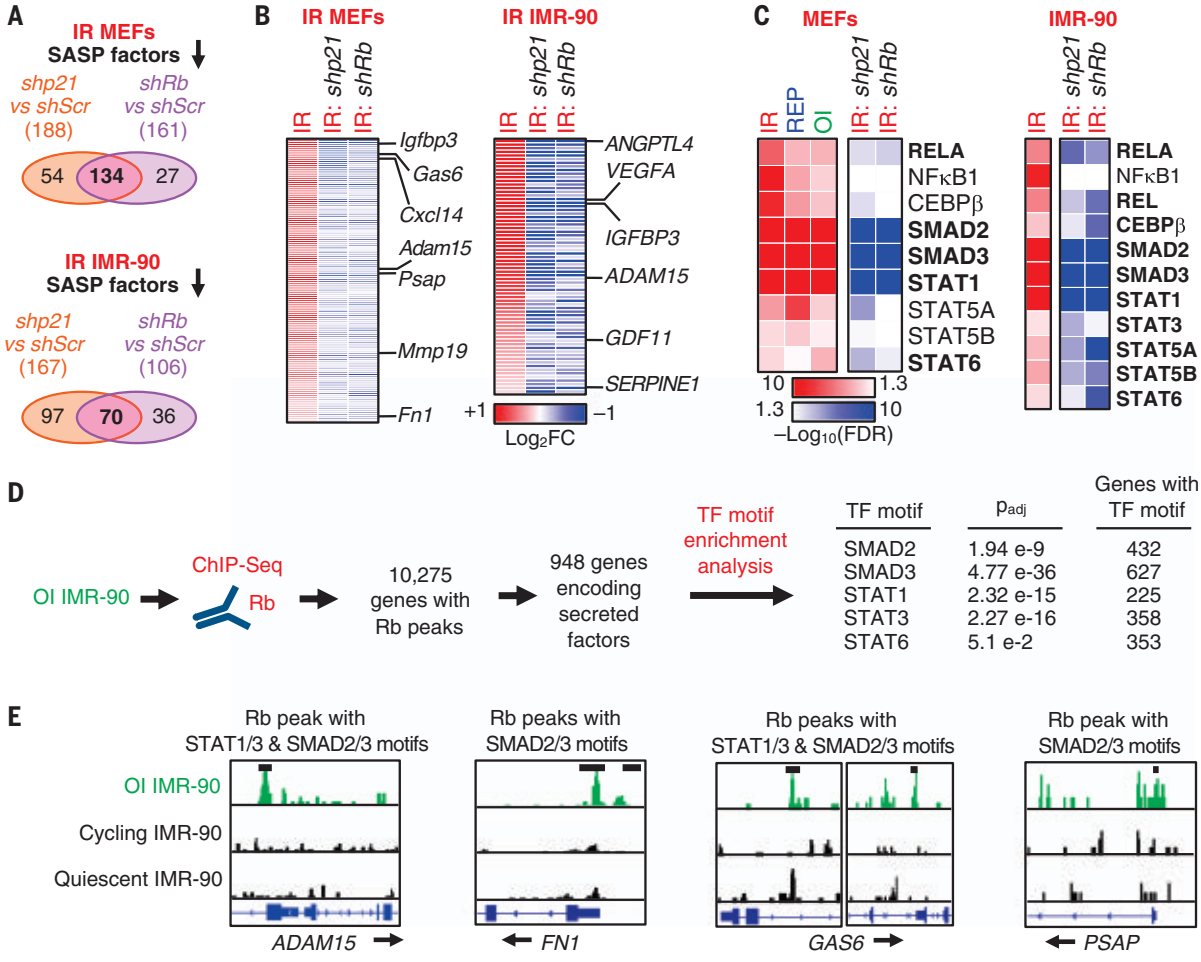


Fig. 1. p21-activated Rb interacts with STAT and SMAD TFs at select gene promoters to establish a bioactive secretome. (**A**) Venn diagrams of RNA-seq data depicting down-regulated SASP factors with depletion of *p21* or *Rb* in the indicated SNCs. (**B**) Heatmaps of commonly down-regulated SASP factors indicated in (A). FC, fold change. (**C**) Identification of TFs that transcriptionally activate genes encoding PASP factors using overrepresentation analyses on RNA-seq data from IR-, REP, OI-senescent MEFs, IR-senescent IMR-90 cells, and their nonsenescent counterparts. Bolded TFs are significantly activated in SNCs and inhibited upon shRNA-mediated depletion of *p21* or *Rb*. FDR, false discovery rate. (**D**) Identification of SASP genes that bind Rb, and TF motif analysis of Rb peaks underlying SFs in OI-senescent IMR-90 cells. (**E**) Representative Rb occupancy plots at PASP genes. Experiments in the above panels were performed once.

Rb appears to enhance the transcriptional activity of these TFs in SNCs to establish the PASP.

Analysis of publicly available Rb ChIP-seq data from OI-senescent IMR-90 cells (*25*) revealed that Rb peaks mapped to the promoter regions of 948 SFs and that these peaks were enriched for binding sites of all TFs that we identified as instrumental in establishing the PASP, with the exception of RELA (Fig. 1D and table S5). Rb peaks mapped to promoter regions of 49 of 172 PASP genes identified in IR IMR-90 cells and associated with TFs critical for establishing the PASP (Fig. 1E and table S5). Most of these promoter regions had no such peaks when IMR-90 cells were cycling or quiescent. Furthermore, SMAD2, SMAD3, STAT1, and STAT6 coimmunoprecipitated Rb from IR-senescent MEFs, and codepletion of SMAD2, SMAD3, STAT1, and STAT6 in IR-senescent MEFs reduced transcription of PASP genes where Rb and these TFs colocalize in promoter regions (fig. S6). Thus, a p21-responsive Rb pool interacts with specific SMAD and STAT TFs at PASP gene promoters to enhance their expression.

p21 simultaneously places cells under proliferative arrest and immunosurveillance

To determine whether the PASP is senescence dependent, we performed RNA-seq on nonsenescent MEFs with high p21 collected 2 or 4 days (d2 or d4) after irradiation (Fig. 2A and fig. S7, A to D). At d2 and d4, IR MEFs upregulated 351 and 450 SFs, respectively, 241 of which were shared with d10 IR MEFs (Fig. 2A and table S6). At d4, IR MEFs depleted for *p21* or *Rb*, lost 235 and 171 of their SFs, respectively, indicating that the PASP is a senescence-independent phenomenon (Fig. 2B and fig. S7, A to D). Eighty-four PASP factors were commonly lost in d4 and d10 IR MEFs when p21 or Rb was depleted, demonstrating that the PASP of non-SNCs becomes an integral part of the SASP as cells advance to a senescent state (Fig. 2B and fig. S7, E and F).

Functional annotation analysis on the 84 shared PASP factors indicated that several traits of SNCs may be p21-Rb dependent, including features involving cell migration and adhesion and the immune system (Fig. 2C and fig. S7G). This raises the possibility that the PASP plays a role in immunosurveillance. To test this idea, we determined the extent to which the PASP affects the migratory behavior of mouse peritoneal immune cells in a transwell system (Fig. 2D). Conditioned medium from d4 nonsenescent (CM-NS) or d10 senescent (CM-S) IR MEFs promoted transwell migration of macrophages, a property that was lost with CM-NS and CM-S from *p21*- or *Rb*-depleted IR MEFs (Fig. 2E). None of the CMs affected lymphocyte migration in this assay

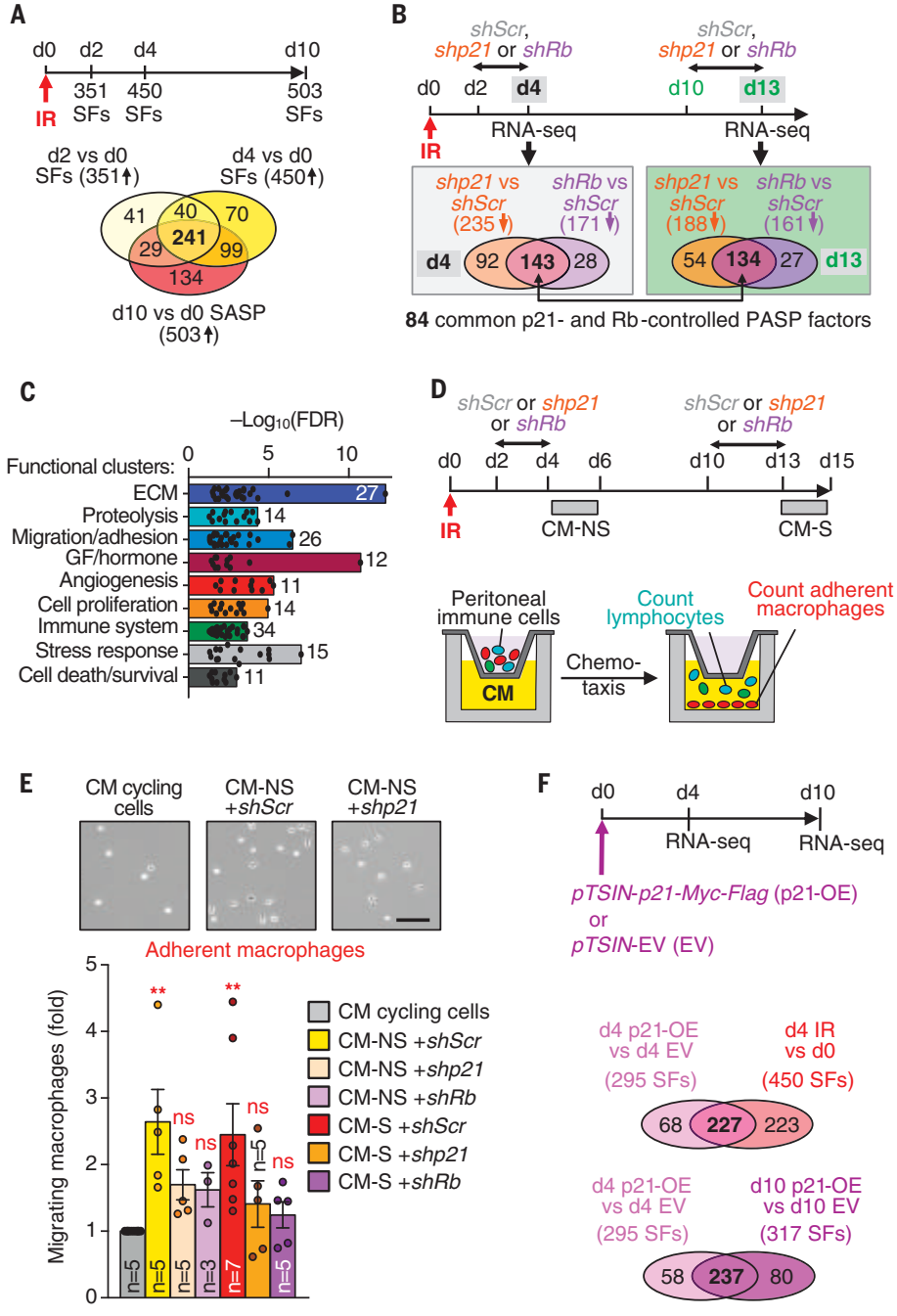


Fig. 2. The p21-activated secretory phenotype places cells under immunosurveillance as they arrest. (**A**) Timeline and Venn diagrams based on RNA-seq depicting significantly up-regulated SFs. (**B**) Timeline and Venn diagrams comparing significantly down-regulated SFs upon *p21* or *Rb* depletion. (**C**) Functional annotation analyses of 84 PASP factors indicated in (B) displaying overrepresented functional clusters. ECM, extracellular matrix; GF, growth factor. (**D**) Schematic of CM production and transwell migration assay of peritoneal immune cells in the presence of CM. (**E**) Representative images and quantitation of adherent macrophages in the bottom transwell chamber. Scale bar: 100 μm. (**F**) Venn diagrams depicting significantly up-regulated PASP factors. Data represent means ± SEM. Experiments in the above panels were performed once except for (E), where data from three experiments were pooled. ns, not significant. $^{**}P < 0.01$. One-way ANOVA with Sidak's correction (E).

(fig. S8A). In a second migration assay, macrophage numbers selectively increased in the peritoneal lavage 4 days after intraperitoneal injection of CM-NS, but not after injection of CM-NS from *p21*- or *Rb*-depleted IR MEFs (fig. S8, B to F). The PASP also stimulated cell movement in a standard scratch assay on cultured MEFs (fig. S8, G and H), indicating that its promigratory properties extend beyond macrophages. Nuclear factor κB (NF-κB) p65 (RELA) appeared to have no role in establishing the PASP or its macrophage-attractant properties (fig. S9 and table S7).

To determine whether the PASP requires an actual senescence-inducing stressor or merely elevated p21 levels, we transduced MEFs with a lentivirus harboring *p21-Myc-Flag* (fig. S10A). p21-overexpressing (p21-OE) MEFs were subject to growth arrest, initially without elevated *p16* and SA–β-galactosidase (β-Gal) activity (d4), and later with these senescence markers (d10) (fig. S10, B to D). At d4, p21-OE MEFs up-regulated 295 SFs, 227 of which were also up-regulated in d4 IR MEFs, indicating that p21 induction is sufficient to yield a PASP (Fig. 2F and table S8). SMAD2, SMAD3, STAT1, and STAT6 coimmunoprecipitated Rb from the chromatin fraction of d2 p21-OE MEFs (fig. S10E), further supporting the contention that p21-induced, hypophosphorylated Rb interacts with STAT and SMAD TFs at select gene promoters to establish the PASP. PASP factors of d4 p21-OE MEFs were largely preserved in d10 p21-OE MEFs (Fig. 2F and table S8). Thus, the PASP becomes an integral part of the SASP as cells senesce.

p21 enforces immunosurveillance through the chemokine CXCL14

Functional annotation analysis on the PASP of d4 p21-OE MEFs suggested that it has biological properties similar to those of the PASP of d4 IR-MEFs (fig. S10F). Indeed, CM from d4 p21-OE MEFs stimulated fibroblast migration in our scratch assay and macrophage migration in our transwell assay and increased local macrophage numbers when intraperitoneally injected in mice (fig. S10, G to N). MEF-derived PASPs consistently included *Cxcl14* (Fig. 1B and table S8), a member of the CXC chemokine family that acts as a chemoattractant for various immune cell types (*26*–*28*). The addition of CXCL14-neutralizing antibodies to CM harvested from d4 p21-OE MEFs ablated stimulation of macrophage migration in our transwell assay, whereas control immunoglobulin G (IgG) did not (Fig. 3A and fig. S11A). Moreover, CM from *Cxcl14*-depleted d4 p21-OE MEFs failed to evoke macrophage migration (Fig. 3B and fig. S11, B and C), further demonstrating that CXCL14 is the key macrophage attractant of the PASP. Complementary experiments in human dermal fibroblasts (HDFs) and human umbilical vein endothelial cells (HUVECs) suggested that the PASP is a common feature of p21 induction and that CXCL14 is a signature PASP component (fig. S12).

To study the PASP phenomenon at the organismal level, we established a transgenic approach in mice that allows for Cre-inducible coexpression of Myc-Flag–tagged p21 and the fluorescent reporter protein tdTomato (Tom) (Fig. 3C). We induced Myc-Flag–tagged p21 and Tom in ~10% of hepatocytes by tail vein

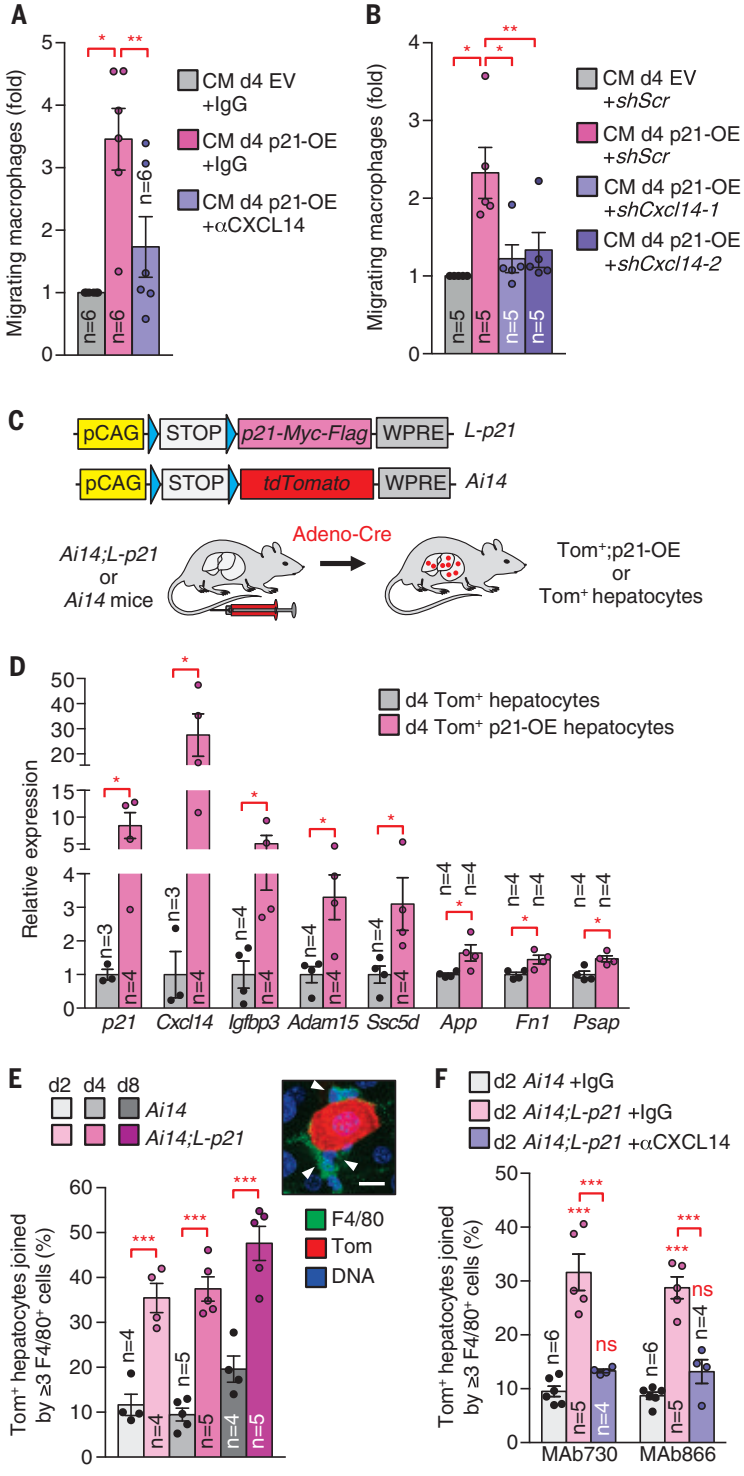


Fig. 3. p21-induced immunosurveillance requires PASP factor CXCL14. (**A**) Transwell migration assay with CM in the presence of CXCL14-neutralizing or IgG antibodies. (**B**) as in (A) but with CM from shRNA-transduced MEFs. (**C**) Schematic of the transgenes used for Cre-inducible expression of Myc-Flag–tagged p21 (*L-p21* for *LoxP/Stop/LoxP-p21*) and tdTomato (*Ai14*) (top) and the experimental setup for p21-OE induction in hepatocytes via Cre-encoding adenovirus injection (bottom). (**D**) RT-qPCR on flow-sorted Tom$^+$ hepatocytes. (**E**) Representative image and quantification of Tom$^+$ hepatocytes joined by ≥3 F4/80$^+$ cells (arrowheads). (**F**) As in (E) but assessing livers from mice treated with CXCL14-neutralizing or IgG control antibodies. Scale bars: 10 μm (E). Data represent means ± SEM. Experiments in the above panels were performed once. ns, not significant. *P < 0.05; **P < 0.01; ***P < 0.001. One-way ANOVA with Sidak's correction [(A), (B), (E), and (F)] and unpaired two-tailed t tests (D).

injection of adeno-Cre virus. Hepatocytes with p21 overexpression (p21-OE) were cell-cycle arrested at d4 after injection and exhibited signs of senescence by d8 after injection, as evidenced by loss of lamin B1 and nuclear extrusion of HMGB1 (fig. S13, A to C). RNA extracted from d4 Tom$^+$ hepatocytes collected by fluorescence-activated cell sorting (FACS) and used for RT-qPCR analysis of PASP factor gene transcripts indicated that p21-OE induces a PASP in vivo (Fig. 3D and fig. S13D). *Cxcl14* was among the up-regulated PASP factors, prompting us to test whether p21-OE hepatocytes attract macrophages. Nearly 40% of p21-OE Tom$^+$ hepatocytes were surrounded by three or more macrophages as early as d2 after adeno-Cre injection versus ~10% of Tom$^+$ hepatocytes without p21-OE (Fig. 3E). Macrophage recruitment to p21-OE hepatocytes was CXCL14 dependent as assessed by injection of antibodies against CXCL14 (Fig. 3F). Lymphocytes were also recruited but later, with B and T cells surrounding p21-OE hepatocytes at d4 and d8, respectively (Fig. 4, A and B). By contrast, p21-OE did not prompt recruitment of natural killer (NK) cells (fig. S13E). The number of p21-OE hepatocytes declined sharply by d8, which coincided with a marked increase in dying p21-OE hepatocytes in the presence of M1-differentiated macrophages in addition to the presence of both CD4$^+$ and CD8$^+$ T lymphocytes (Fig. 4, C to F, and figs. S13F and S14, A and B). Administration of a CD8α-neutralizing antibody fully prevented the observed decline in d8 p21-OE hepatocytes (fig. S14, C to G), indicating that their elimination is mediated by cytotoxic T cells.

We performed a comparative analysis for overexpression of *p16* (p16-OE), a more selective cyclin-dependent kinase (CDK) inhibitor, which, unlike p21, only targets G_1-CDK activity. At d4, p16-OE MEFs were characterized by growth inhibition, normal *p21* levels, and a secretome of 197 factors, 183 of which overlap with the PASP of d4 p21-OE MEFs (fig. S15, A to F, and table S8). Pathway enrichment analyses on the p16-associated secretory phenotype suggested a high degree of similarity in biological properties with the PASP, although the number of immune-system–related annotations was markedly reduced (fig. S15, G and H). CM of d4 p16-OE MEFs failed to promote migration of macrophages in our transwell assay, which correlated with a lack of *Cxcl14* induction (fig. S15I). Likewise, using the same transgenic approach as used for p21-OE in mice, we found that p16-OE in hepatocytes triggered cell-cycle arrest but not immunosurveillance, which coincided with a lack of p21 and PASP factor induction, including *Cxcl14* (fig. S16). Corresponding analyses of MEFs overexpressing p27, a CDK inhibitor that enables cell-cycle withdrawal during terminal differentiation,

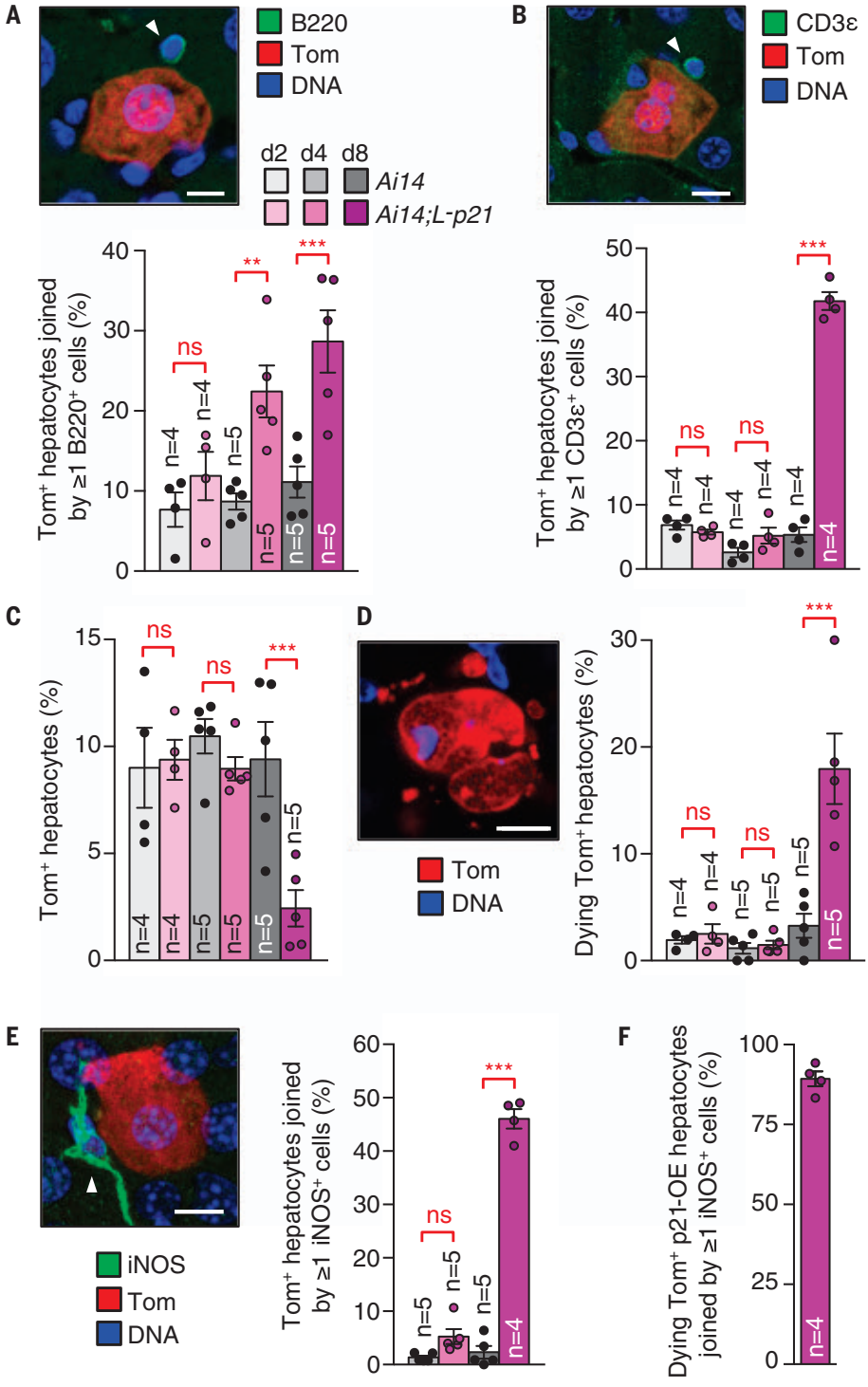


Fig. 4. Hepatocytes under surveillance die upon macrophage differentiation and lymphocyte recruitment. (**A**) Representative image and quantification of Tom$^+$ hepatocytes associated with ≥1 B220$^+$ cells (arrowhead). (**B**) Representative image and quantification of Tom$^+$ hepatocytes associated with ≥1 CD3ε$^+$ cells (arrowhead). (**C**) Proportion of Tom$^+$ and healthy (not dying) hepatocytes. (**D**) Representative image and quantification of dying Tom$^+$ hepatocytes. (**E**) Representative image and quantification of Tom$^+$ hepatocytes associated with ≥1 iNOS$^+$ cells (arrowhead). (**F**) Quantitation of dying p21-OE Tom$^+$ hepatocytes associated with ≥1 iNOS$^+$ cells. Scale bars: 10 μm [(A), (B), (D), and (E)]. Mice used were on a C57BL/6×129Sv mixed genetic background. Data represent means ± SEM. Experiments in the above panels were performed once. ns, not significant. $^{**}P < 0.01$; $^{***}P < 0.001$. One-way ANOVA with Sidak's correction [(A) to (E)].

revealed that coordinated induction of growth arrest and immunosurveillance is a distinctive feature of p21 (fig. S17 and table S8).

Oncogene-induced p21 triggers immunosurveillance

We sought to critically test the physiological relevance of p21-dependent immunosurveillance in a cancer-related context. To this end, we adapted our transgenic approach for co-induction of Tom and p21 in hepatocytes by replacing *p21* with *KRAS*^{G12V} (Fig. 5A), an oncoprotein that can induce p21 via mitogenic stress (*4*). Indeed, ~25% of d4 Tom$^+$ KRAS^{G12V} hepatocytes had elevated levels of p21 (Fig. 5B and fig. S18A). These hepatocytes attracted macrophages, whereas those that failed to induce p21 did not (Fig. 5C). Use of a newly generated *p21* conditional knockout strain conclusively demonstrated that d4 Tom$^+$ KRAS^{G12V} hepatocytes recruit macrophages in a p21-dependent manner (Fig. 5, A to C, and fig. S18B). Furthermore, d4 Tom$^+$ KRAS^{G12V} hepatocytes in which *Rb* was conditionally knocked out retained p21 induction but nevertheless failed to attract macrophages, validating cell culture experiments indicating that p21 places cells under immunosurveillance in an Rb-dependent fashion (Fig. 5, A to C, and fig. S18B).

Hepatocytes from d4 Tom$^+$ KRAS^{G12V} mice had a PASP, which they lost with conditional inactivation of *p21* (Fig. 6A). The PASP included *Cxcl14*, explaining why d4 Tom$^+$ KRAS^{G12V} hepatocytes attract macrophages and their counterparts lacking p21 do not. Tom$^+$ hepatocyte numbers remained largely unchanged at d4, d12, and d28 after induction when KRAS^{G12V} was absent (Fig. 6B), but progressively declined because of cell death when KRAS^{G12V} was coexpressed (Fig. 6, B and C). However, no such decline occurred when *p21* was inactivated upon KRAS^{G12V} induction. This was not due to compensatory cell proliferation because *p21* inactivation had no impact on the mitotic index of Tom$^+$ KRAS^{G12V} hepatocytes (fig. S18, C and D). Consistent with p21-dependent cell elimination, Tom$^+$ KRAS^{G12V} hepatocytes with high p21 levels gradually decreased from d4 to d28 (Fig. 5B). At d12, Tom$^+$ KRAS^{G12V} hepatocytes were surrounded by M1 macrophages and T lymphocytes, whereas their d4 counterparts were not (Fig. 6, D and E). Thus, p21-mediated macrophage recruitment represents an essential first step toward cell elimination upon KRAS^{G12V} induction, which then appears to be executed by subsequently recruited T lymphocytes. This conclusion is supported by the earlier observation that clearance of hepatocytes expressing oncogenic NRAS depends on monocytes and macrophages and on CD4 T cells (*4*).

Regardless of whether *p21* was intact or inactivated, Tom$^+$ KRAS^{G12V} hepatocytes hardly proliferated and showed signs of cellular

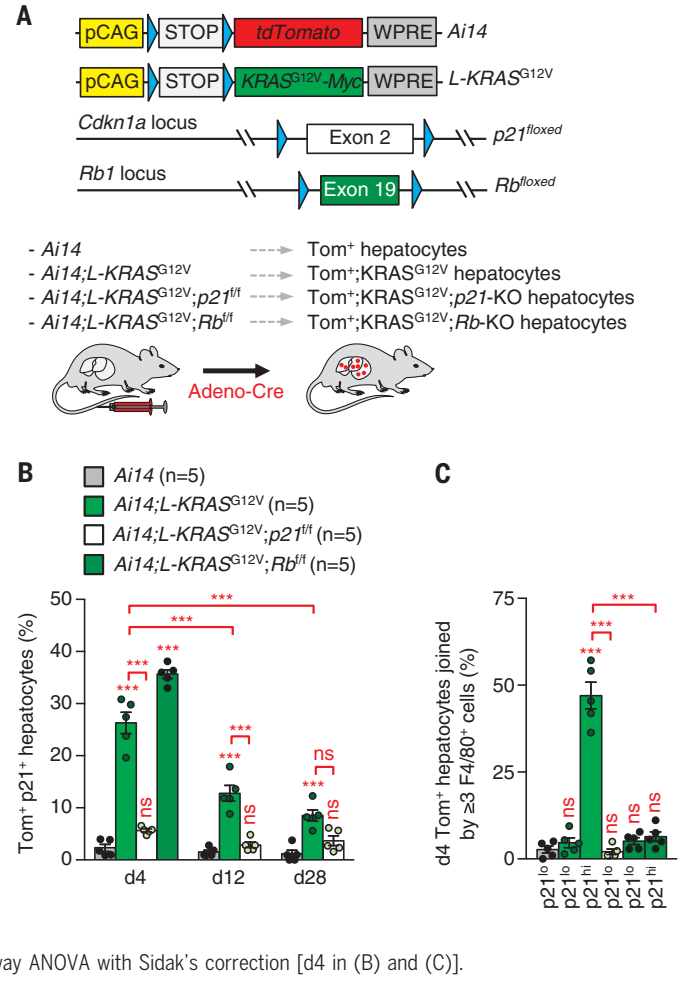


Fig. 5. p21 induced by oncogenic RAS places cells under immunosurveillance. (**A**) (Top) Schematic representation of *L-KRAS*^{G12V} and *Ai14* transgenes, and *p21*- and *Rb*-conditional knockout alleles. Blue triangles denote LoxP sites. (Bottom) Schematic of the experimental design. (**B**) Proportion of Tom$^+$ p21$^+$ hepatocytes among Tom$^+$ hepatocytes at indicated days after adeno-Cre injection. (**C**) Quantification of Tom$^+$ hepatocytes joined by ≥3 F4/80$^+$ macrophages. p21^{hi}, cells with elevated p21 staining; p21^{lo}, cells with baseline or background level p21 staining. Mice used were on a C57BL/6×129Sv mixed genetic background. Data represent means ± SEM. Experiments in the above panels were performed once. ns, not significant. ***$P < 0.001$. Two-way ANOVA with Sidak's correction [d12 and d28 in (B)], one-way ANOVA with Sidak's correction [d4 in (B) and (C)].

senescence from d12 on (fig. S18, C to F). However, small clusters of Tom$^+$ KRAS^{G12V} hepatocytes were observed in d28 livers with much higher frequency when *p21* was inactivated (Fig. 6F). Hepatocytes within these clusters were cycling at a markedly higher rate than corresponding hepatocytes located in isolation (Fig. 6G). Thus, p21-dependent immunoclearance of cells that experience oncogenic stress constitutes an important first line of defense against neoplastic growth.

p21 engages macrophages to set a biological timer that controls cell fate

Stress-inducing oncogenic point mutations are irreparable, but many cellular stresses are transient or repairable. To determine whether stressed cells that recuperate and normalize p21 cease to produce a PASP and are released from immunosurveillance, we produced MEFs containing a lentiviral construct that allows for doxycycline (dox)–inducible expression of *p21-Myc-Flag*. These MEFs stopped proliferating within 2 days after dox administration, but were fully capable of resuming the cell cycle when p21 levels normalized upon dox withdrawal (Fig. 7, A and B, and fig. S19, A to C). CM harvested from d2 p21-OE MEFs stimulated fibroblast migration in the scratch assay and macrophage migration in the transwell assay with peritoneal immune cells (Fig. 7C and fig. S19, D and E). By contrast, CM prepared from MEFs that had been on dox for 2 days followed by 4 days off dox had no impact on cell migratory properties in these same assays. *Cxcl14* expression followed the promigratory properties of p21-OE CM, with reduced *Cxcl14* expression upon withdrawal corresponding to a more generalized collapse in PASP gene expression (Fig. 7D). Suppression of E2F target genes, proliferative arrest, induction of PASP genes, and chemoattraction of macrophages all occurred within 24 hours after induction of p21 (fig. S19, F to J). Thus, the brake on cell-cycle progression and PASP-mediated immune surveillance occur simultaneously and rapidly.

To determine the time allocated to damaged cells to recuperate and avert elimination by immune cells under physiological conditions and to define the underlying timer mechanism, we created transgenic mice in which p21 could be temporally overexpressed in hepatocytes by sequential administration of adeno-Cre injection and dox (Fig. 8A). Macrophages surrounding hepatocytes overexpressing p21 for up to 4 days promptly withdrew upon normalization of p21 (Fig. 8, B and C, and fig. S20, A and B). By contrast, when p21 levels were normalized at day 6, macrophages surrounding hepatocytes failed to disengage. At this stage, hepatocytes were adjoined not only by macrophages that had undergone M1 activation, but also lymphocytes that had been primed for target cell elimination (Fig. 8, D to F). At d6, p21-OE hepatocytes were not yet senescent, although some entry into the senescent state occurred during the 2-day off period (fig. S20, C to E). Thus, p21 induction sets a timeframe for stress repair or adaptation that is defined by the time it takes for the immune system to transition from a cell-surveillance to a cell-clearance mode.

Discussion

In response to a wide variety of stress stimuli, p53 induces p21 to halt the cell cycle as one of the hundreds of downstream p53 transcriptional targets. As such, p21 provides time for cells to recuperate through various repair and adaptation mechanisms, or if the damage is too severe, for induction of apoptosis or for cellular senescence. Our studies revealed that p21 also responds to cellular stressors through a cell-nonautonomous mechanism by placing cells under immunosurveillance and that p21 does so concomitantly with halting cell-cycle progression. These findings extend earlier evidence linking p21 inactivation to immune system deficiencies and predisposition to systemic autoimmunity (*29*–*31*). In probing the mechanism, we discovered that the pool of hypophosphorylated Rb that is created in response to p21 induction binds to chromatin to not only establish cell-cycle arrest but also to cooperate with select SMAD and STAT TFs to create a bioactive secretome with diverse biological functions, including immunosurveillance.

We found that CXCL14 within this secretome is instrumental in attracting macrophages to cells with elevated p21. By recruiting macrophages, p21 sets a biological timer that allows for a period of repair or adaptation of about 4 days (at least for hepatocytes), which expires when macrophages polarize toward an M1 phenotype and lymphocytes are recruited to mount a cytotoxic T cell response. Therefore, the p21-mediated killing of stressed cells by the leveraging of the immune system represents an important complementary mechanism to p53-mediated apoptosis. Hepatocytes that induce p21 in response to a well-established oncogenic stressor that drives hepatocellular carcinoma are eliminated via this mechanism, underscoring its physiological relevance. Thus, the first line of tumor-protective immunosurveillance appears to already occur at the earliest stages of cellular transformation,

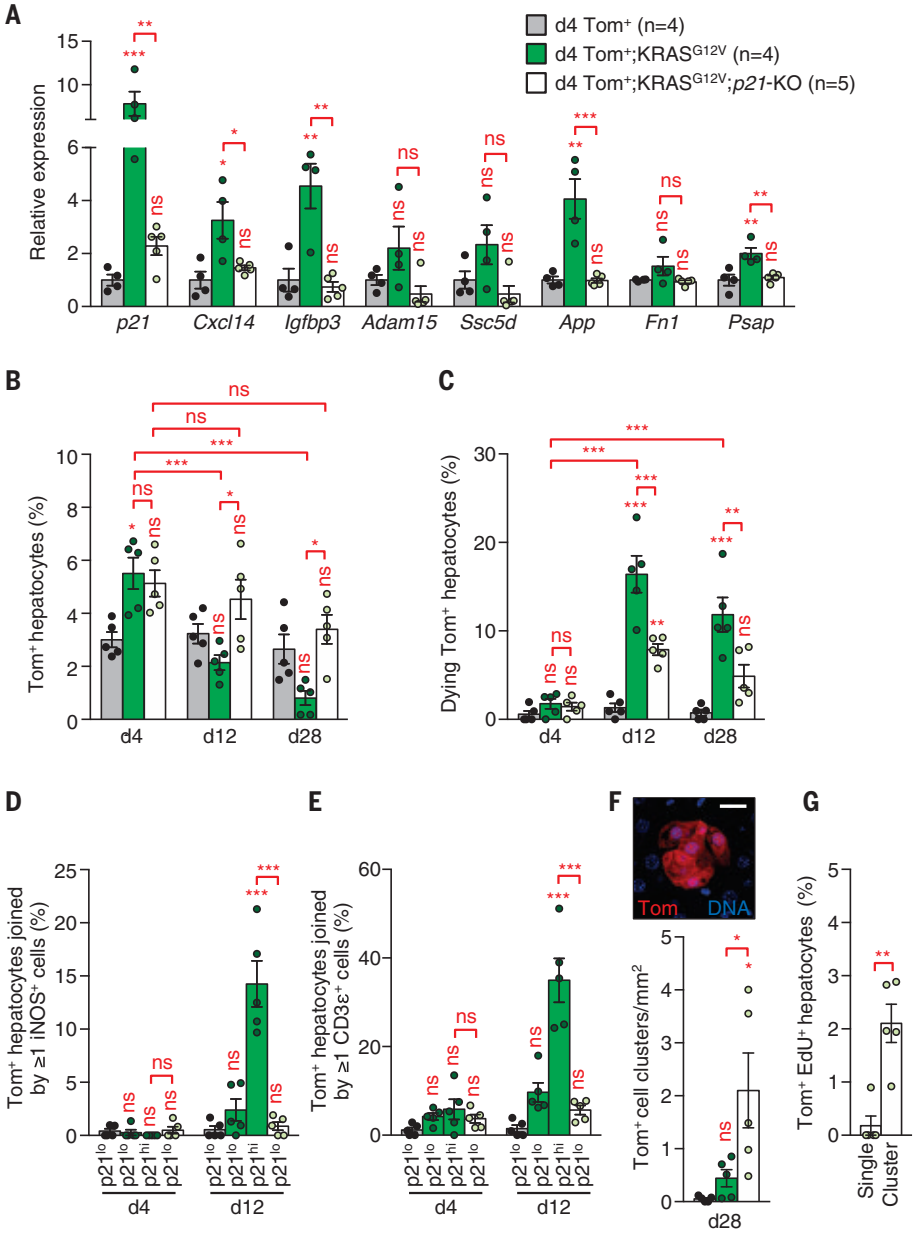


Fig. 6. p21-dependent immunoclearance protects against oncogenic growth. (**A**) RT-qPCR on flow-sorted Tom$^+$ hepatocytes. (**B**) Proportion of hepatocytes that are Tom$^+$ and appear healthy (not dying). (**C**) Quantification of dying Tom$^+$ hepatocytes. (**D**) Quantification of Tom$^+$ hepatocytes joined by ≥1 iNOS$^+$ cells. p21^{hi}, cells with elevated p21 staining; p21^{lo}, cells with baseline or background level p21 staining. (**E**) As in (D) but for hepatocytes with ≥1 CD3ε$^+$ cells. (**F**) Representative image and quantitation of Tom$^+$ hepatocyte clusters. (**G**) Proportion of Tom$^+$ EdU$^+$ hepatocytes in- or outside Tom$^+$ clusters. Scale bar: 20 μm (F). Mice used were on a C57BL/6×129Sv mixed genetic background. Data represent means ± SEM. Experiments in the above panels were performed once. ns, not significant. *P < 0.05; **P < 0.01; ***P < 0.001. Two-way ANOVA with Sidak's correction [(d12 and d28 in (B) and (C)], one-way ANOVA with Sidak's correction [(A) and (D) to (F) and d4 in (B) and (C)] or unpaired two-tailed t test (G).

complementing established tumor-protective actions of immune cells based on the expression of tumor-specific neoantigens (*32*). As a transcriptional regulator, p53 has been implicated in diverse aspects of immunity, including tumor-related immune responses (*33*). The same holds true for Rb, raising the interesting question of the extent to which the roles of these two tumor suppressors are p21 dependent.

Our comparative analysis of p21-OE with p16-OE and p27-OE suggests that coordinated cell-cycle arrest and transcriptional activation of select genes is a common feature of CDK inhibitors that act through Rb hypophosphorylation. We observed substantial overlap between genes activated by the three inhibitors, but many genes were also specifically induced. On the surface, this argues against the idea that Rb dephosphorylation is the underlying driver. However, hypophosphorylated Rb appears to be a collection of Rb molecules that are monophosphorylated at one of 14 known phosphorylation sites, with each variant having a characteristic binding-partner profile and a distinctive transcriptional output (*34*). It is therefore conceivable that p21, p16, and p27 create distinctive collections of monophosphorylated Rb molecules, which, through differences in transcription factor binding preferences, activate distinct target genes, thereby yielding secretomes of different complexities and biological properties. Our observation that p16-OE does not induce CXCL14 explains why these cells were not subject to immune clearance. Consistent with this finding, cells undergoing senescence through interferon-γ– and tumor necrosis factor–mediated induction of p16 are not subject to elimination (*35*).

In conclusion, we demonstrate that p21, through Rb hypophosphorylation, induces the macrophage attractant CXCL14 to place cells under immunosurveillance as part of an immediate early response to cellular stress that also includes cell-cycle arrest. Our study highlights the role of CXCL14, but the PASP contains several other immune-modulatory factors, including interleukin-7 (IL-7) and IL-34. Both these interleukins induce M1 macrophage differentiation (*36*, *37*), a phenotypic change that we show coincides with the recruitment of cytotoxic T cells to p21-expressing hepatocytes. A greater understanding of IL-7 and IL-34 and other PASP factors in stress-related immunosurveillance could inform innovative therapeutic opportunities for the treatment of cancers, including cancers lacking p53 activity. The observation that low *CXCL14* expression correlates with poor clinical outcome in various human cancers, including cervical, colorectal, endometrial, and head and neck cancers (*38*), underscores the relevance of such future efforts.

Materials and methods

Mouse strains

L-KRAS^{G12V} mice were generated from KH2 ES (C57BL/6×129Sv) cells according to previously described methods (*39*) using a modified *pBS31* vector. Briefly, the tetracycline-inducible promoter and the SV40 polyA signal in the *pBS31* were replaced by a CAG promoter-FRT-loxP-flanked STOP cassette (*LoxP-STOP-LoxP, L*) and *WPRE-bGH-polyA* (*WPRE-pA*) from *Ai9* (Addgene, #22799), respectively. The FRT site after the CAG promoter was deleted using site-directed mutagenesis and a multiple cloning site (MCS) was added between *L* and *WPRE-pA*.

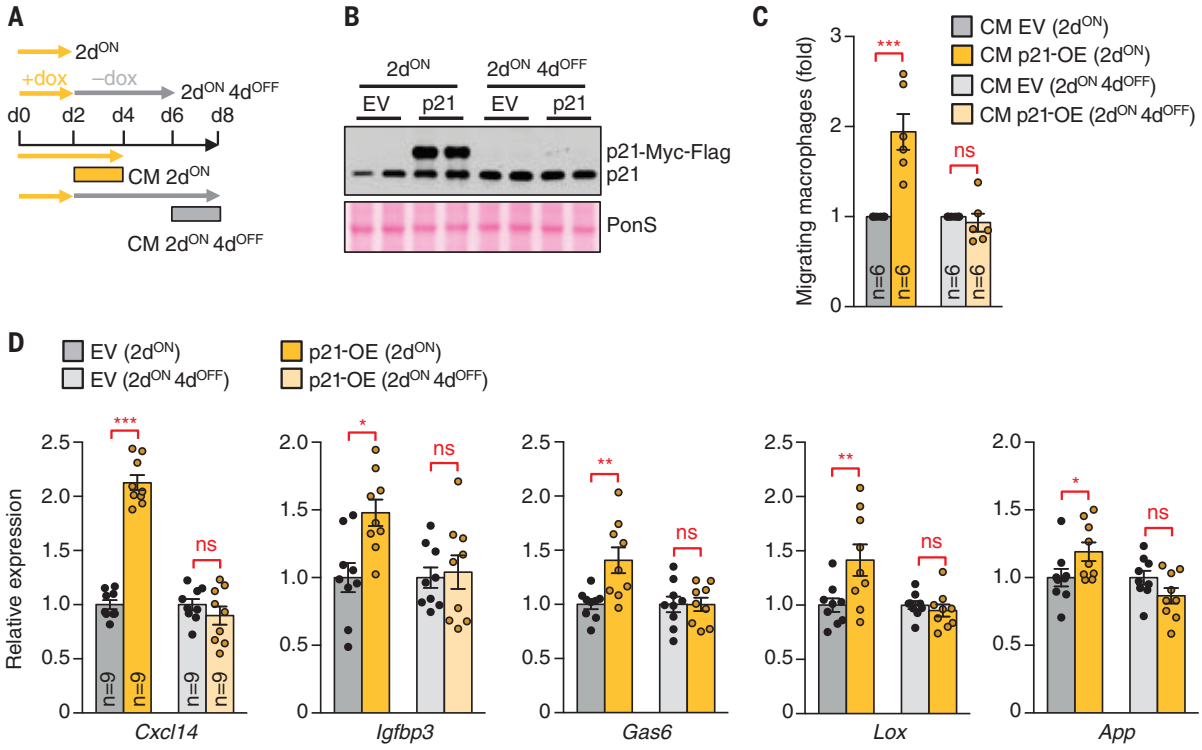


Fig. 7. Cells normalizing p21 cease to produce a PASP and are released from immunosurveillance. (**A**) Schematic overview of CM preparations from dox-inducible p21-OE MEFs. (**B**) Representative immunoblot for p21. PonS served as loading control. (**C**) Transwell macrophage migration with CM from indicated MEFs. (**D**) RT-qPCR of the indicated MEFs. Data represent means ± SEM. The blot shown in (B) is representative of two independent experiments, and data shown in (C) and (D) were pooled from two and three experiments, respectively. ns, not significant. *P < 0.05; **P < 0.01; ***P < 0.001. Two-way ANOVA with Sidak's correction [(C) and (D)].

The Myc-tagged human *KRAS*^{G12V} was amplified from *pBABE-KRAS*^{G12V}-puro (Addgene, #9052) and inserted to the MCS. The resultant *pBS31-CAG-L-KRAS*^{G12V}*-WPRE-pA* plasmid was co-electroporated with an FLP expression plasmid into KH2 ES cells and selected clones with Cre-inducible *KRAS*^{G12V} expression were used to generate *L-KRAS*^{G12V} mice according to standard procedures. The same strategy was used to generate *L-p21* mice or *L-p16* mice using Myc-Flag-tagged cDNAs for mouse *p21* or mouse *p16* obtained from Origene (#MR227529 or #MR227284, respectively). Obtained founder mice (C57BL/6×129Sv background) were backcrossed to C57BL/6 at least twice before use in breeding for experimentation. To generate *iL-p21* transgenic mice on a pure C57BL/6 background, the following targeting construct was cloned: *pTRE2-LoxP-STOP-LoxP(LSL)-p21-Myc-Flag-WPRE-pA* using the *TRE2* promoter and *LSL* from the *Ai139* transgene (Addgene, #114426) and *p21-Myc-Flag* from the *L-p21* transgene (Origene, #MR227529) as described above. Homology arms spanning 968 bp at the 5′-end and 937 bp at the 3′-end flanked by sgRNA target sites were used to target the construct into the *Col1a1* locus of C57BL/6NHsd (Envigo) zygotes using CRISPR-Cas9-mediated gene editing with *Cas9* mRNA (Trilink Biotechnologies, #L-7606) and *Col1a1*-specific sgRNA 5′-GAGGTTCATGAGCCCTCAAA-3′. Obtained founder mice were backcrossed to C57BL/6 once before use for experimentation. To generate *p21*^{floxed} mice on a pure C57BL/6 background, a targeting vector containing *p21* exon 2 flanked by LoxP sites and homology arms spanning 861 bp at the 5′-end and 819 bp at the 3′-end flanked by sgRNA target sites (5′ sgRNA: 5′-TCTTGGTGATTAACTCCATC-3′ and 3′ sgRNA: 5′-CCATAGGCGTGGGACCTCGT-3′) was cloned. The resultant targeting vector was used to target the construct into the *p21* locus of C57BL/6NHsd (Envigo) zygotes using CRISPR-Cas9-mediated gene editing with *Cas9* mRNA (Trilink Biotechnologies, #L-7606) (*40*). Obtained founder mice were crossed to C57BL/6 at least once before use for experimentation. *Rb*^{floxed} mice (#026563), *Ai14* transgenic animals (#007914), and *Ai139* transgenic mice (#030219) were purchased from The Jackson Laboratory. The following cohorts were generated for experimentation in this study: *Ai14/+* and *Ai14/+ L-KRAS*^{G12V}*/+* (C57BL/6×129SV mixed background; fig. S2), *Ai14/+* and *Ai14/+ L-p21/+* (C57BL/6×129SV mixed background; Figs. 3 and 4, figs. S13 and S14), *Ai14/+* and *Ai14/+ L-p16/+* (C57BL/6×129SV mixed background; fig. S16), *Ai14/+* and *Ai14/+ L-KRAS*^{G12V}*/+* and *Ai14/+ L-KRAS*^{G12V}*/+ p21*$^{floxed/floxed}$ and *Ai14/+ L-KRAS*^{G12V}*/+ Rb*$^{floxed/floxed}$ (C57BL/6×129SV mixed background; Figs. 5 and 6 and fig. S18), *Ai139/+* and *Ai139/+ iL-p21/+* (C57BL/6 pure background; Fig. 8 and fig. S20). Mice were aged until 4 to 6 months of age before use for experimentation unless otherwise noted. Experimental procedures involving laboratory mice were reviewed and approved by the Institutional Animal Care and Use Committee of the Mayo Clinic.

Cell culture

Mouse embryonic fibroblasts (MEFs) were generated and cultured as described previously with each line being derived from a separate C57BL/6 E13.5 embryo containing *INK-ATTAC* (*12*, *41*). These lines were expanded at 3% oxygen and used for experiments between passage (P)3 and P6. IMR-90 cells were purchased from ATCC (#CCL-186) at P10 and cultured in the same medium as used for MEFs. IMR-90 cells were used for experimentation between P14 and P18. Human dermal fibroblasts (HDFs) were generated from human foreskin of young, healthy donors (2 days to 13 years of age) by the Biochemical Genetics Laboratory at the Mayo Clinic. Each line was derived from a separate donor. HDFs were cultured in the same medium as used for MEFs and used for experimentation between P5 and P8. HUVECs were purchased from ATCC (#PCS-100-013) and were

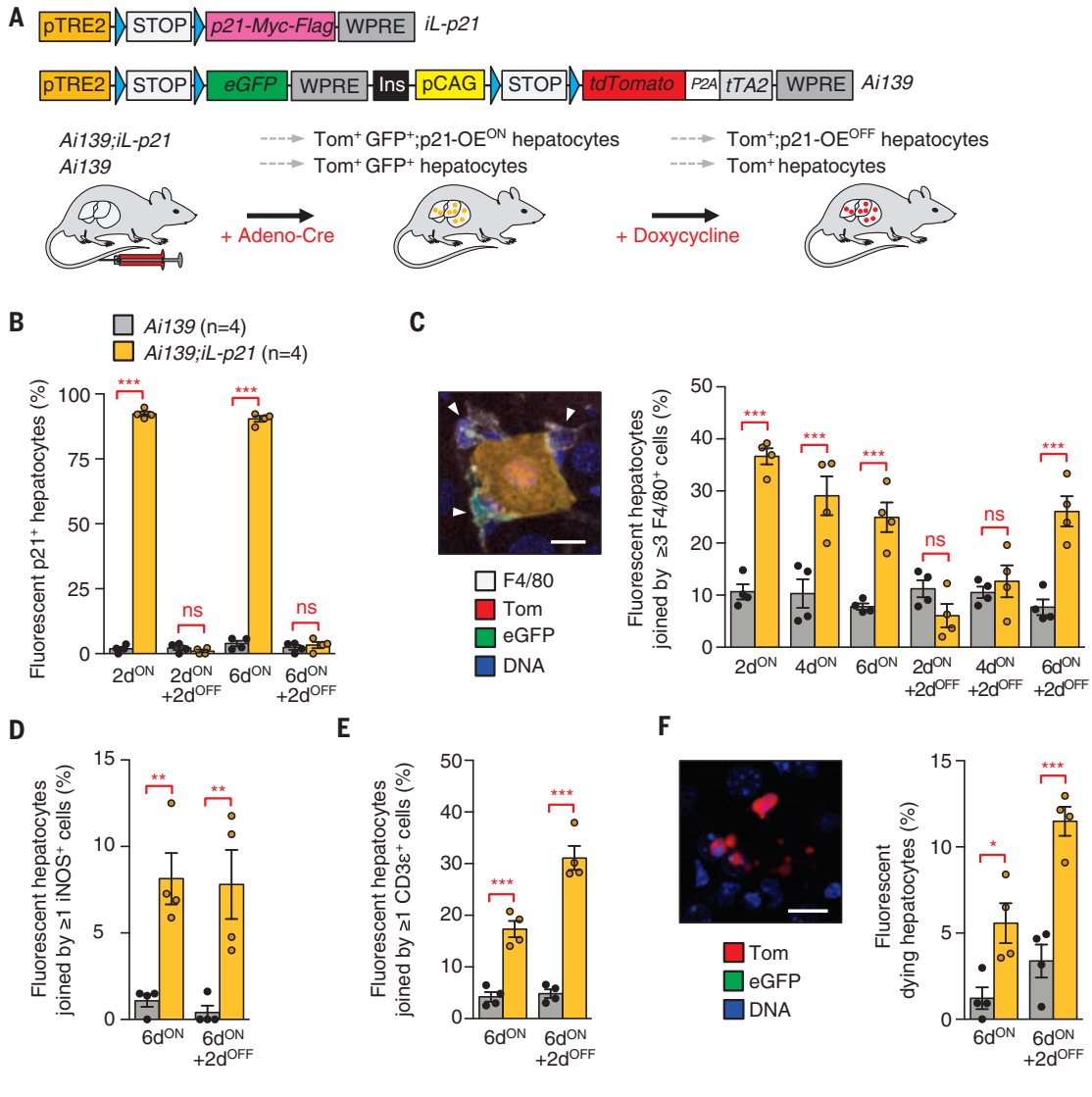


Fig. 8. p21 places cells under immunosurveillance to establish a timer mechanism that controls cell fate. (**A**) (Top) Schematic representation of the transgenes for dox-inducible repression of Myc-Flag–tagged p21 (*iL-p21* for inducible *LoxP/Stop/LoxP-p21*) and monitoring of ectopic p21 expression status via fluorescent marker proteins (*Ai139*). Blue triangles denote LoxP sites. (Bottom) Schematic of the experimental design with fluorescent markers for transgenic p21 expression and repression indicated. (**B**) Rates of p21 overexpression (p21^{+}) among hepatocytes that were positive for Tom and eGFP (enhanced green fluorescent protein) in the absence of dox (p21-OE "ON") or only Tom after dox administration (p21-OE "OFF"). (**C**) Representative image of a p21-OE hepatocyte surrounded by three macrophages (arrowheads), and quantification of fluorescent hepatocytes joined by ≥3 F4/80^{+} macrophages. (**D**) Assessment of fluorescent hepatocytes associated with ≥1 iNOS^{+} cells. (**E**) As in (D) but assessing cells with ≥1 CD3ε^{+} cells. (**F**) Representative image of a 6d^{ON}+2d^{OFF} dying hepatocyte and quantification of death rates among fluorescent hepatocytes. Scale bars: 10 μm [(C) and (F)]. Mice used were on a C57BL/6 pure genetic background. Data represent means ± SEM. Experiments in the above panels were performed once. ns, not significant. *P < 0.05; **P < 0.01; ***P < 0.001. Two-way ANOVA with Sidak's correction [(B) to (F)].

cultured in vascular cell basal medium (ATCC, #PCS-100-030) supplemented with endothelial growth factors (Endothelial Cell Growth Kit-VEGF, ATCC, #PCS-100-041). HUVECs were used for experimentation at P3 to P5 after being received.

Generation of senescent and nonsenescent MEFs

For H3K27Ac-ChIP-seq experiments, two or three independent MEF lines were generated and induced to senesce via irradiation (IR), serial passaging (REP) or KRAS^{G12V}-overexpression (OI). For identification of IR-induced SASEs, we established the following three MEF cultures from each independent MEF line: proliferating P3 MEFs (to derive C1 MEFs); P6 MEFs exposed to 10 Gy γ-radiation (^{137}Caesium source) and cultured for 2 days (to derive C2 MEFs); and P6 MEFs exposed to 10 Gy γ-radiation and cultured for 10 days (to derive IR-senescent MEFs). For identification of REP-induced SASEs, we prepared two MEF cultures from each independent MEF line: proliferating P3 MEFs (to derive C1 MEFs) and P10 MEFs cultured at 20% oxygen between P4 and P10 (to derive REP-senescent MEFs). To identify SASEs in OI-induced senescent MEFs, cells were infected with a *KRAS*^{G12V}-containing lentivirus (prepared using the *pLenti-PGK-ER-KRAS*^{G12V} from Addgene #35635), selected with 250 μg/ml of hygromycin B (EMD Millipore, #400052) and then harvested (to derive C1 MEFs) or treated with 200 nM 4-hydroxytamoxifen (4′-OHT, 1:50,000 from stock in ethanol, Sigma H7904) to induce KRAS^{G12V} for 2 days (to derive C2 MEFs) or 10 days (to derive OI-induced senescent MEFs). IR-, REP- and OI-induced senescent MEFs were enriched by sterile FACS as previously described using a BD FACSAria 4-laser digital flow cytometer with FACSDiva v8.0.1 software with 488 nm laser (*42*). Sorted cells were pelleted, resuspended in fresh culture medium, counted, and used for ChIP-seq and RNA extraction. Small amounts of the sorted cells were reseeded to assess the proportion of cell that was senescent. Samples with ~70% or more SNCs were used for H3K27ac-ChIP-seq experiments. C1 and C2 MEFs cultures were also subjected to FACS but here fractions devoid of SNCs were collected. For all other experiments involving REP-induced SNCs, SNCs were prepared as described above. FACS-enriched SNCs were cultured for at least 24 hours before further use. OI-induced senescent MEFs were also prepared as described above, but instead of the lentiviral KRAS^{G12V} expression system we used MEFs derived from

L-KRAS^{G12V} mice. These MEFs were infected with *pTSIN-Cre-PGK-puro2* lentivirus (*43*) to induce KRAS^{G12V} expression. These MEFs were then cultured for 10 days and subject to FACS enrichment of SNCs (the first 2 days in medium containing 2 μg/ml of puromycin).

Generation of IR-senescent and control IMR-90 cells

H3K27Ac-ChIP-seq experiments and matched RNA-sequencing experiments were conducted in triplicate using three technical replicates. IMR-90 cells were expanded at 3% oxygen and used for experiments at P18. For identification of IR-induced SASEs, we established the following three cultures from each of the replicates: proliferating P18 IMR-90 cells (to derive control 1 (C1) cells); P18 IMR-90 cells exposed to 10 Gy γ-radiation (^{137}Caesium source) and cultured for 2 days (to derive control 2 (C2) cells); and P18 IMR-90 cells exposed to 10 Gy γ-radiation and cultured for 10 days (to derive IR-senescent IMR-90 cells). Cells were trypsinized and reseeded to assess the proportion of SNCs Samples with >80% IR-SNCs were used for H3K27Ac ChIP-seq experiments.

ChIP-seq analyses and super-enhancer identification in cultured cells

FACS-enriched MEF or IMR-90 suspensions were pelleted, resuspended in medium, and counted. A total of 2×10^5 to 1×10^6 cells were fixed with 1% paraformaldehyde (PFA) for 10 min and then subjected to ChIP-seq as previously described using a rabbit anti-H3K27Ac antibody (Abcam, ab4729, Lot GR150367) (*44*). Chromatin immunoprecipitation–sequencing (ChIP-seq) libraries were prepared from 1-5 ng of precipitated chromatin or input DNA using the Ovation ultralow DR Multiplex kit (NuGEN) or the ThruPLEX DNA-seq Kit V2 (Rubicon Genomics). ChIP enrichment was validated in library DNAs by performing quantitative PCR in the indicated genomic loci using following primers: mouse *mPabpc1-TSS* (F): 5′-ATCCCACAGCTTGTGGCGGG-3′; (R): 5′-TCTCGCCATCGGTCGCTCTC-3′; mIntergenic (F): 5′-CCT-GCTGCCTTGTCTCTCTC-3′; (R): 5′-ATGGCCTAGGGATTCCAGCA-3′. The ChIP-seq libraries were sequenced to 51 bp from both ends on an Illumina HiSeq 2000 or HiSeq 4000 instrument in the Mayo Clinic Medical Genomics Core Facility. Fastq files of pair-end reads were mapped with Bowtie 1.1.2 using parameters -k 1 -m 1 -e 70 -l 51–best to the reference genome as previously described (mm10 for mouse, hg19 for human) (*23*, *45*). We used MACS 1.4.2 to identify peaks for each sample against the background using a *P*-value cutoff of 10^{-5} (*46*). All other parameters were left at default. To identify super-enhancers, neighboring peaks were first stitched together to create a single region capturing these signals as a whole. Peaks occurring within 12.5 kb from each other were combined into stitched enhancers while excluding regions that were within ±2000 bps from any transcription start site (TSS) (*23*). These stitched enhancers were then ranked by background-subtracted ChIP-seq occupancy ascendingly, and the occupancy was plotted in the unit of reads per million per base pair. From the plot, we identified the point where occupancy started increasing faster by first scaling the *x*- and *y*-axes into [0, 1] and then finding the point where a line with a slope of 1 was tangential to the curve. Occupancy increased slowly below but rapidly above this point. The stitched enhancers above this point were defined as super-enhancers. All the above procedures were performed using ROSE (*21*, *23*). In order to determine differential binding for SE between treatment and control samples, we first merged SE regions from all samples into a set of merged regions covering all super-enhancer regions in all samples. Tag counts at each merged region were then extracted and differential analysis on the tag counts were performed using R package DESeq2 1.10.1 using the same settings as described below (see RNA sequencing) (*47*). SASEs were defined as super-enhancers with lfcMLE (unshrunk $\log_2$ fold change produced by DESeq2) in tag counts ≥0.3 for both senescent versus proliferating (C1) and senescent versus induced, nonsenescent (C2). Super-enhancers were assigned to genes within ±50 kb of the super-enhancer by calculating the distance between either end of each super-enhancer and TSS of each gene (*21*). Only super-enhancers ±50 kb from at least one TSS were considered in downstream analyses. For downstream validation, only SASE-controlled genes that were differentially expressed with false discovery rate (FDR)<0.05 in at least two of three senescence mechanisms were considered. BigWig files of H3K27Ac occupancy profiles were generated using deepTools 3.1.0 (*48*) by first normalizing each ChIP-seq sample and its matching input to cpm (counts per million mapped reads) and then subtracting the input signal from each ChIP sample. H3K27Ac occupancy plots were generated via Integrative Genomics Viewer (IGV) (*49*). To identify Rb peaks at promoters of SFs, published Rb ChIP-seq data from OI-senescent, quiescent and nonsenescent IMR-90 cells were analyzed (GSE19899) (*25*). Peaks were annotated to genes within 50 kb from either end of any peak. The peak sequences of SASP genes associated to any Rb peak with 2.5-kb padding from each end were used as input to MEME-ChIP (*50*) to detect enriched motifs using the HOCOMOCO database (*51*). We used FIMO to locate occurrences of motifs in each input sequence (*52*).

ChIP on senescent liver cells

FACS-enriched Tom$^+$ cell suspensions from *Ai14;L-KRAS*^{G12V} or *Ai14* control livers (see below) were pelleted, resuspended in medium, and counted. A total of $1\text{-}4\times10^5$ cells were fixed with 1% PFA for 10 min and then subjected to H3K27Ac-ChIP using a rabbit anti-H3K27Ac antibody (Abcam, ab4729, Lot GR150367) or rabbit, IgG (Millipore, #12-370) according to the manufacturers protocol (Active Motif, #53084). Precipitated chromatin or input DNA was subjected to quantitative PCR in the indicated genomic regions in the SASE of the *p21* locus using primers indicated in table S9.

RNA isolation and RT-qPCR

MEFs or IMR-90 cells, or flow-sorted liver cells were lysed in RLT buffer supplemented with β-mercaptoethanol according to the RNA extraction protocol. RNA extraction (Qiagen, RNeasy Mini kit, #74104, or RNeasy Micro kit, #74004), cDNA synthesis (Invitrogen, SuperScript III First-Strand Synthesis, #18080051), and quantitative polymerase chain reaction analysis (Applied Biosystem, SYBR Green Real-Time PCR Master Mix, #4309155) were performed according to manufacturer's instructions. The on-column DNase digestion step was avoided during the RNA extraction procedure unless RNA was used for RNA-sequencing purposes. Primers were optimized via cDNA dilution series. *Tbp* (*TBP* in human) was used as a reference gene for RT-qPCR in mouse and human samples. Primer sequences are listed in table S9.

RNA sequencing

Equal amounts of high-quality RNA (100-200 ng) were subjected to library preparation using the TruSeq RNA Library Prep Kit v2 (Illumina, #RS-122-2001) according to the manufacturer's instructions. Libraries were sequenced following Illumina's standard protocol using the Illumina cBot and HiSeq 3000/4000 PE Cluster Kit. Flow cells were sequenced as 100×2 paired-end reads on an Illumina HiSeq 4000 using HiSeq 3000/4000 sequencing kit and HCS 3.3.20 collection software. Base-calling was performed using Illumina's RTA 2.5.2 software. RNA sequencing was performed at the Mayo Clinic Center for Individualized Medicine Medical Genomics Facility (Mayo Clinic, Rochester, Minnesota). Fastq files of pair-end RNA-seq reads were aligned with Tophat 2.0.14 to the reference genome (mm10 for mouse, hg19 for human) using Bowtie2 2.2.6 with default parameters (*51*, *53*, *54*). Gene level counts were obtained using FeatureCounts 1.4.6 from the SubRead package with gene models from corresponding UCSC annotation packages (*55*). Differential expression analysis was performed using R package DESeq2 1.10.1 after removing genes with average raw counts less than 10 (*47*). During the DESeq2 analysis thresholding on Cook's distance for outliers and independent filtering were turned off so that all genes passed to DESeq2 were assigned

P-values for significance of differential expression. Genes with FDR<0.05 were considered significantly differentially expressed. Hierarchical clustering of samples was performed using DESeq2-normalized counts with 1−Pearson correlation as distance and average linkage using R function hclust. Gene Set Enrichment Analysis (GSEA) was performed as previously described against mouse genesets from Enrichment map using gene lists ranked by lfcMLE, which was the unshrunk $\log_2$ fold change produced by DESeq2, in descending order (*56*, *57*). Functional annotation analyses were performed via String database v11 focusing on GO BP annotations, KEGG pathways and Reactome pathways with FDR<0.05 (*58*). Overrepresentation analysis for transcription regulatory targets of individual TFs was performed using the Fisher's exact test method for selected gene lists against the mouse gene sets from ENCODE and MSigDB collections (*57*, *59*). Mouse TF targets were mapped to human orthologs using MGI's Vertebrate Homology database and used for overrepresentation analyses in human datasets. Putative SASP factor genes were extracted from Gene Ontology Consortium (*Mus musculus* MGI and *Homo sapiens* GO Annotations EBI) and QuickGO database for the annotation GO:0005615 "Extracellular Space" (*60*–*62*). Gene lists from both reference databases were merged resulting in the identification of 1845 and 3513 factors for mouse or human, respectively. Heatmaps were generated with Morpheus, Broad Institute (https://software.broadinstitute.org/morpheus). For gene expression heatmaps based on RNA-seq data, lfcMLE values and $-\log_{10}$ of FDR values were used.

Adenovirus injection into mice and isolation of liver cells

To generate in vivo OI-senescent liver cells, we used 4-month-old *Ai14;L-KRAS*^{G12V} or *Ai14* control mice injected with 100 µl of 0.9% NaCl containing 10^9 pfu adeno-Cre-EGFP virus (University of Iowa, Vector Labs, #VVC-U of Iowa-1174) into the tail vein. Eight days post-injection, livers were harvested and the perivenous half of the left lateral lobe was fixed with 4% PFA in PBS for 2 hours and incubated in 30% sucrose overnight. These livers were embedded in OCT (Sakura, #4583) and used for cryosectioning and confocal imaging. To assess proliferation rates in these mice, 50 mg per kg of EdU (5-ethynyl-2′-deoxyuridine, Carbosynth, #NE08701) was IP injected on days 6 and 7 post-adeno-Cre injection for a total of 48 hours before euthanasia of mice. EdU staining was performed on cyrosections with the same kit and protocol used in vitro (see below). To isolate Tom$^+$ liver cells, livers of *Ai14;L-KRAS*^{G12V} or *Ai14* control mice 8 days post-injection were perfused with collagenase as previously described (*63*, *64*). Because the parenchymal fraction of *Ai14;L-KRAS*^{G12V} was not viable, the non-parenchymal fraction was subjected to FACS as described above with appropriate lasers and filters. For in vivo p21-OE and p16-OE studies, *Ai14;L-p21* or *Ai14;L-p16* or *Ai14* control mice were injected with 100 µl of 0.9% NaCl containing 10^8 pfu adeno-Cre-EGFP virus into the tail vein. Livers were harvested and fixed as described above 2, 4, or 8 days post-injection. To assess proliferation rates, these mice were injected intraperitoneally with 50 mg per kg of EdU on days 2 and 3 after injection of adeno-Cre-EGFP virus and euthanized 48 hours after the first EdU injection. For in vivo KRAS^{G12V}-OE studies, *Ai14;L-KRAS*^{G12V}, *Ai14;L-KRAS*^{G12V} *p21*$^{floxed/floxed}$, *Ai14;L-KRAS*^{G12V} *Rb*$^{floxed/floxed}$ or *Ai14* control mice were injected with 100 µl of 0.9% NaCl containing 2.5×10^7 pfu adeno-Cre-EGFP virus into the tail vein. Livers were harvested and fixed as described above, 4, 12, or 28 days post-injection. To assess proliferation rates in these mice, 50 mg per kg of EdU was injected IP on 2 days and 1 day for a total of 48 hours before euthanasia of mice. To isolate Tom$^+$ hepatocytes for expression analyses, livers were perfused with collagenase and the parenchymal fraction was subjected to FACS as described above. For in vivo inducible p21-OE studies, *Ai139;iL-p21* or *Ai139* control mice were injected with 100 µl of 0.9% NaCl containing 10^8 pfu adeno-Cre-EGFP virus into the tail vein. At indicated timepoints ("ON"), livers were harvested and fixed as described above. To suppress p21-OE ("OFF"), mice were treated with doxycycline (Letco, #690902) at 100 mg per kg in water via gavage every 24 hours (for a total of 48 hours) until euthanasia and liver collection.

DNA isolation and PCR for recombined conditional alleles

Livers of *A14i;L-KRAS*^{G12V}, *A14i;L-KRAS*^{G12V}*;p21*$^{floxed/floxed}$ or *A14i;L-KRAS*^{G12V}*;Rb*$^{floxed/floxed}$ mice that received 2.5×10^7 pfu adeno-Cre virus (containing ~5% Tom$^+$ hepatocytes) or did not receive virus were flash frozen and stored at −80°C. These livers were homogenized via mortar and pestle and DNA was isolated through phenol-chloroform extraction. PCR analysis of *p21* exon 2 was performed using the following primers: (F) 5′-GTATCCCAAAGTCCAGGGCACT-3′ and (R) 5′-TGCCAAGGGGAAGGACATCATT-3′ generating 1446 bp, 1549 bp, and 609 bp products for the wild-type, unrecombined-floxed, and recombined-floxed alleles, respectively. PCR analysis of *Rb* exon 19 was performed as described before (*65*) using the following primers Rb18 (F) 5′-GGCGTGTGCATCAATG-3′ and Rb212 (R) 5′-GAAAGGAAAGTCAGGGACATTGGG-3′ generating 698 bp, 746 bp, and 260 bp products for the wild-type, unrecombined-floxed, and recombined-floxed alleles, respectively.

Neutralizing antibody experiments in mice

To deplete CD8$^+$ T cells, *Ai14;L-p21* and *Ai14* mice were IP injected with 500 µg of rat anti-CD8α antibody (clone 53-6.7, BioXcell, #BE0004-1) in 200 µl PBS or 200 µl PBS (as control) each day for 3 consecutive days and again on day 6. On day 7, mice were injected with 100 µl of 0.9% NaCl containing 10^8 pfu adeno-Cre-EGFP virus into the tail vein. On d12, mice were IP injected once more with anti-CD8α antibody or PBS, mice were euthanized and livers and spleens were collected at d15 (corresponding to d8 post-adeno-Cre injection). Spleens were processed freshly to isolate cells for flow cytometry. Spleens were dissociated between two frosted slides. The cell suspension was then filtered through a 70-µm filter and spun at 400 *g* for 5 min. Red blood cells were removed via ACK lysis for 8 min on ice. Tubes were filled with PBS, centrifuged, and resuspended. Total cell numbers were then counted. For flow cytometry assessments, 10^6 cells were used for antibody staining using the following antibodies: hamster FITC-conjugated anti-TCRβ (Tonbo Biosciences, #35-5961, 1:500), PerCP-conjugated rat anti-CD4 (BioLegend, #100538, 1:500), violetFluor 450-conjugated rat anti-CD8α (clone 2.43, Tonbo biosciences, #75-1886, 1:500), and Ghost Dye Red 780 cell viability dye (Tonbo biosciences, #13-0865, 1:1000). Total CD4$^+$ or CD8α$^+$ T cells were calculated using flow cytometry quantifications and the previously noted total cell numbers per spleens. To neutralize CXCL14, *Ai14;L-p21* and *Ai14* mice were IP injected with the following antibodies in 200 µl of PBS: 500 µg of rat anti-CXCL14 antibody (R&D Systems, #MAb730), 500 µg of mouse anti-CXCL14 antibody (R&D Systems, #MAb866), 500 µg of mouse IgG2a isotype control (BioXcell, #BE0085 as control for Mab730), or 500 µg of rat IgG2b isotype control (BioXcell, #BE0090 as control for MAb866). The following day, mice were again injected with antibody and were also injected with 10^8 pfu adeno-Cre virus in 100 of µl 0.9% NaCl intravenously as described above. The following day, antibody injection was repeated once more. Mice were euthanized and livers were collected the next day (d3, corresponds to d2 post-adeno-Cre injection).

Cryosectioning and immunofluorescence on liver tissue

OCT-embedded livers were sectioned using a Cryostat (CM 1900, Leica) to generate 20-µm-thick frozen sections. Sections were washed with PBS and permeabilized with 0.5% Triton-X-100 for 20 min. Sections were blocked with 5% BSA/PBS for 1 hour and subsequently incubated overnight with primary antibodies rabbit anti-F4/80 (Cell Signaling, #70076; 1:250), FITC-conjugated rat anti-B220/CD45R (BD BioSciences; #553088; 1:50), FITC-conjugated rat anti-NKp46/CD335 (Biolegend, #580756; 1:50),

rabbit anti-CD3ε (Cell Signaling, #99940; 1:50), biotin-conjugated rabbit anti-CD4 (BioLegend, #100508, 1:50; in combination with FITC-conjugated streptavidin, BioLegend, #405201, 1:100), rabbit anti-CD8α (Cell Signaling, #98941, 1:20), rabbit anti-iNOS (Abcam, ab15323, 1:100), rabbit anti-lamin B1 (Abcam, ab16048, 1:500) or rabbit anti-HMGB1 (Abcam, ab18256, 1:500), rabbit anti-p21 (Abcam, ab188224, 1:100 or 1:250), rabbit anti-Myc-tag (Cell Signaling, #2272, 1:100), mouse anti-Myc-tag (Cell Signaling, #2276, 1:100; in combination with Alexa Fluor 647-conjugated goat anti-mouse IgG2a secondary antibody, Invitrogen, #A21241, 1:100), rabbit anti-phospho-Histone H3 (Ser10) (pHH3, Millipore, #06-570, 1:250), or Alexa Fluor 488-conjugated rat anti-F4/80- (Bio-Rad, #MCA497A488T, 1:100; used for coimmunofluorescence in combination with rabbit anti-p21 staining) diluted in 5% BSA/PBS. Secondary antibodies used were Alexa Fluor 488-conjugated goat anti-rabbit IgG (Invitrogen, #A11034; 1:250) or Alexa Fluor 647-conjugated goat anti-rabbit IgG (Invitrogen, #A21244; 1:100). These antibodies were incubated for 3 hours. Secondary antibody incubations were avoided if the primary antibody was conjugated to FITC or Alexa Fluor fluorophores. Washes between incubations were performed in PBS (three 5-min washes). Cells were counterstained with Hoechst. A confocal laser-scanning microscope (LSM 880; Zeiss) on an Axio Observer Z1 inverted microscope with spectral detectors (32ch 2PMT GaAsP; Zeiss) and a water-immersion lens (C-Apochromate 40X/1.2 NA Korr. FCS; Zeiss) were used to capture *z*-stack images with 2-μm step size (F4/80, iNOS, NKp46, CD3ε, CD4, CD8α, and B220 stainings). The percentage of lamin B1$^+$ nuclei was determined as the percentage of Tom$^+$ hepatocytes with lamin B1-staining versus Tom$^+$ hepatocytes without lamin B1 staining. At least 50 hepatocytes or two sections were counted. For HMGB1 staining, the localization of nuclear versus cytoplasmic staining was examined per Tom$^+$ hepatocyte and percentage of Tom$^+$ hepatocytes with nuclear HMGB1 (N>C) was determined compared to Tom$^+$ hepatocytes with loss of nuclear HMGB1 and gain of cytoplasmic staining (N<C). At least 50 hepatocytes or two sections were counted. To determine the proportion of p21-induced hepatocytes, the percentage of Tom$^+$ hepatocytes with nuclear p21-staining versus Tom$^+$ hepatocytes without nuclear p21 were quantified. At least 100 hepatocytes or two sections were counted. Similar analyses were performed to quantify Myc-tag-induced hepatocytes of *Ai14;L-p21* mice. To determine the proportion of Myc-tag-induced *Ai14;L-KRAS*^{G12V} hepatocytes, the percentage of Tom$^+$ hepatocytes with Myc-tag-staining at the plasma membrane versus Tom$^+$ hepatocytes without Myc-tag staining were quantified. To count the number of macrophages/Kupffer cells, B cells, T cells, or NK cells associated per Tom$^+$ hepatocyte, the number of F4/80$^+$ cells, B220$^+$, CD3ε$^+$, or NKp46$^+$ cells, respectively, immediately adjacent to Tom$^+$ hepatocytes was counted. At least 100 hepatocytes or two sections were counted. Similar quantifications were performed for the M1 macrophage marker iNOS and T cell subset markers CD4 and CD8α. To assess the proportion of Tom$^+$ hepatocytes actively progressing through the cell cycle, Tom$^+$ hepatocytes with nuclear pHH3 staining versus Tom$^+$ hepatocytes without pHH3 signal were quantified. Cells with pHH3 staining were subdivided into Tom$^+$ pHH3$^+$ before nuclear envelop breakdown as determined via Hoechst signal (considered G_2 cells) and after nuclear envelop breakdown (considered mitotic cells). To determine the percentage of Tom$^+$ hepatocytes, at least 400 hepatocytes were scored and the percentage of Tom$^+$ versus Tom$^-$ hepatocytes (as determined by nuclear and cellular shape) were determined. To assess the number of dying hepatocytes, at least 100 Tom$^+$ hepatocytes were examined for cellular health and cells with overtly fragmented cytoplasm were considered as dying. Tom$^+$ hepatocyte clusters were defined as three or more Tom$^+$ hepatocytes being immediately adjacent, whereas Tom$^+$ single hepatocytes were assessed when having no other Tom$^+$ hepatocyte immediately adjacent. To quantify Tom$^+$ hepatocyte clusters, large tile images were captured, assessed for the number of Tom$^+$ hepatocyte clusters and normalized to the area of the tile image. Three sections were analyzed and averaged. For all quantifications involving *Ai139;iL-p21* or *Ai139* mice, similar staining regiments and quantifications were performed, but with the following modifications. In samples without dox ("ON") Tom$^+$ eGFP$^+$ hepatocytes were selected for quantification, whereas in the presence of dox ("OFF") Tom$^+$ hepatocytes were selected. At least 50 Tom$^+$ hepatocytes were examined.

Immunostaining and confocal microscopy

For p21 or 53BP1 immunostaining, flow-sorted MEFs were seeded on 10-well chambered slides (HTC supercured, Thermo Fisher Scientific, #30966S Black) at 2000 cells per well. The following day, cells were fixed in PBS/4% PFA for 15 min, permeabilized in PBS/0.2% Triton X-100 for 15 min and blocked in PBS/5% BSA for 1 hour. Primary antibodies mouse anti-p21 (Santa Cruz, sc-53870; 1:200) or rabbit anti-53BP1 (Novus Biological, #NB100-305; 1:200) were diluted in PBS/5% BSA and subsequently incubated with primary antibodies overnight and secondary antibodies (goat anti-rabbit AlexaFluor488, Invitrogen, #A11034; 1:250) for 3 hours. Washes between incubations were performed in PBS (three 5-min washes). Cells were counterstained with Hoechst and the percentage of p21$^+$ nuclei was determined. For 53BP1 staining, the number of clearly visible 53BP1 foci per cell was counted and percentage of 53BP1$^+$ cells with more than one focus was determined. At least 100 cells or 50 cells per sample were counted for p21 or 53BP1 staining, respectively. A confocal laser-scanning microscope (LSM 880; Zeiss) on an Axio Observer Z1 inverted microscope with spectral detectors (32ch 2PMT GaAsP; Zeiss) and a water-immersion lens (C-Apochromate 40X/1.2 NA Korr. FCS; Zeiss) were used to capture images.

Plasmid constructs

Short hairpin RNA (shRNA) oligo sequences were obtained from the RNAi Consortium (TRC, Broad Institute) and cloned into *pLKO.1* vector (Addgene, #10878) according to a corresponding Addgene protocol (*66*). Four to five shRNAs per gene were tested for their knockdown potential and the two most efficient shRNAs were used in experiments. The non-targeting TRC2 shRNA (referred to as scrambled shRNA *shScr*, Sigma-Aldrich, #SCH202) was used as a negative control. For shRNA sequences see table S9. The Myc-Flag-tagged cDNA for mouse *p21* was obtained from Origene (#MR227529) and subcloned into the lentiviral *pTSIN-PGK-puro2* backbone (*43*, *67*) or dox-inducible *pTRIPZ-PKG-puro* backbone (modified from GE Dharmacon) (*68*). Similarly, the Myc-Flag-tagged cDNAs for mouse *p16* (Origene, #MR227284) and mouse *p27* (Origene, #MR201957) were also subcloned into the lentiviral *pTSIN-PGK-puro2* backbone.

Lentivirus production and cell transduction

Lentiviral particles were produced in HEK-293T cells using Lipofectamine 2000 (Invitrogen, #11668) and appropriate helper plasmids: *pLP1*, *pLP2*, *VSV-G* (*pLKO.1* vectors and *pLenti* vectors), *VSV-G* and *pHR-CMV8.9* (for *pTSIN* vectors) or trans-lentiviral packaging mix (GE Dharmacon, #TLP4606) (for *pTRIPZ* vectors). After 48 hours, virus supernatant was harvested by filtration of HEK-293T supernatant through a 0.45-μm syringe filter. Virus was frozen at −80°C in small aliquots and freshly thawed for each infection cycle.

SA-β-Gal staining

MEFs and IMR-90 cells were seeded on 10-well chambered slides (HTC supercured, Thermo Fisher Scientific, #30966S Black) at 2000 cells per well. Flow-sorted cells were fixed the next day and stained. To assess senescence induction kinetics after irradiation or gene overexpression or gene knockdown, cells were irradiated with 10 Gy or infected twice with appropriate virus supernatants. At indicated times, cells were fixed and stained for SA-β-Gal activity according to manufacturer's protocol (Cell Signaling, #9860S). MEFs were stained for 24 hours, whereas human cells were stained

for 12 hours. Cells were counterstained with Hoechst and the percentage of SA-β-Gal$^+$ cells was then determined. At least 100 cells per sample were counted. To determine the proportion of SA-β-Gal$^+$ hepatocytes in adeno-Cre induced livers, 8-μm-thick cryosections were cut and stained similarly to described previously (*69*). Briefly, sections were fixed for 10 min according to manufacturer's protocol (Cell Signaling, #9860S) and staining was performed for 14 hours. Sections were counterstained with Hoechst. At least 200 hepatocytes (as determined by cell and nuclear shape) were examined for SA-β-Gal$^+$ staining.

Growth curves

Growth curves were generated using senescent MEFs as well as their respective proliferating controls (P5 nonirradiated for IR, P3 for REP, *pLenti-PGK-ER- KRAS*^{G12V}-infected, ethanol-treated cells for OI). At d0, flow-sorted cells were plated in a 12-well plate at a density of 25,000 cells per well in duplicates. At day 4, subconfluent cultures were trypsinized, counted and re-seeded at 25,000 cells per well. Counting was repeated at day 7. Duplicate measures were averaged and cumulative cell number was calculated according to the following formula (*70*) $T_x = T_{x-1} * N_x / N_0$, where T is the cumulative cell number, x the passage number, N_x the counted cell number at passage x, and N_0 the initially seeded cell number. For growth curves of p21-OE or p16-OE MEFs, P3 cells were infected with *pTSIN* empty, *pTSIN-p21-Myc-Flag* or *pTSIN-p16-Myc-Flag* on two consecutive days. The next day (day 3) cells were trypsinized, counted and reseeded in six-well plates at 100,000 cells per well (three wells per condition). Cells were counted every 24 hours until day 6. In parallel, cells were selected with puromycin and reseeded at day 7, when counting was continued.

EdU incorporation assay

Sorted senescent and non-SNCs were seeded on 10-well chambered slides at 2000 cells per well. The next day medium was replaced with medium containing 1 μM EdU (5-ethynyl-2′-deoxyuridine, 1:10,000 dilution, stock in DMSO) and cells were allowed to incorporate EdU for 48 hours. Cells were then fixed and subjected to EdU staining according to the manufacturer's instructions (Thermo Scientific, Click-iT Plus EdU Alexa Fluor 555 Imaging Kit, #C10637). To assess DNA reduplication after knockdown of SASE-controlled genes, senescent MEFs were seeded on 10-well chambered slides at 2000 cells per well and infected with shRNA-containing virus on the 2 following consecutive days. Forty-eight hours after the first infection, medium was replaced with medium containing 1 μM EdU for 48 hours. Four days after the first infection, cells were fixed and subjected to EdU staining. To assess proliferation of irradiated, nonsenescent, P3 MEFs were seeded at 2000 cells per well. The next day, cells were irradiated with 10 Gy. Two days post-IR, EdU was added for 24 hours, or cells were infected with shRNA-virus on two consecutive days. On day 4 post-IR, EdU was added for 24 hours. To assess proliferation of p21-OE human cells or p27-OE MEFs, cycling cells were infected with appropriate virus supernatants for 2 consecutive days as described above, selected for the next 48 hours with 2 μg/ml of puromycin. At day 4, cells were reseeded at 2000 cells per well and EdU was allowed to be incorporated for 24 hours. For inducible p21-overexpression, stably virus-infected cells were reseeded at 2000 cells per well and 4 μg/ml of dox was added the next day. At indicated days, EdU was added for 24 hours, except for short p21-OE induction experiments represented in fig. S19G where EdU was allowed to be incorporated for 12 hours. To quantify the EdU$^+$ cell fraction, cells were counterstained with Hoechst and percentage of EdU$^+$ cells was determined. At least 100 cells were counted.

Immunoblot analysis and coimmunoprecipitation

Coimmunoprecipitations and immunoblot analyses were performed as previously described (*71*). Subcellular fractionation for coimmunoprecipitations on chromatin fractions was performed using the Subcellular Protein Fractionation Kit (Thermo Scientific, #78840) according to the manufacturer's instructions. The primary antibodies used were as follows: mouse anti-p21 (Santa Cruz, sc-53870; 1:8000 used for both mouse and human samples), rabbit anti-Myc-tag (Cell Signaling, #2272; 1:1000); rabbit anti-Rb (Abcam, ab181616; 1:2000), rabbit anti-STAT1 (Abcam, ab92506; 1:1000), rabbit anti-STAT6 (Cell Signaling, #5397; 1:1000), rabbit anti-SMAD2 (Cell Signaling, #5339, 1:1000), rabbit anti-SMAD3 (Cell Signaling, #9513; 1:1000), and mouse anti-p27 (BD Biosciences, #610242, 1:1000). All antibodies were detected with secondary HRP-conjugated goat anti-mouse or anti-rabbit antibodies (Jackson Immunoresearch; 1:10,000). PonS staining (0.2% w/v in 5% glacial acetic acid, Sigma-Aldrich, #P3504) served as a loading control. Immunoblots are representative of at least two independent experiments.

Conditioned medium

To generate CM from IR-induced cells, MEFs were seeded in T75 flasks at low density. The following day, cells were exposed to 10 Gy IR. Two days post-IR, cells were infected with shRNA virus as described above. At day 4 post-IR, these cells as well as cycling control cells of similar density were washed twice and 5 ml of culture medium as added. After 48 hours of conditioning, CM was harvested, filtered through a 0.2-μm syringe filter, and stored in small aliquots at −80°C. To generate CM from IR-SNCs, cells 10 days after IR were used and treated the same way. To produce CM from gene overexpressing MEFs, cells were seeded in T75 flasks, infected with appropriate virus supernatants on the next two consecutive days. Cells were selected with puromycin until d4 or d10 post-infection. Cells were again washed twice before of 5 ml of culture medium was added. CM was harvested as described above. For inducible *pTRIPZ-p21-Flag-Myc* overexpression, 4 μg/ml of dox was added to cells for 48 hours, then cells were washed and subjected to conditioning in the presence of dox, or cells were washed twice immediately and regular culture medium was added. These cells were washed twice a day to remove any residual dox and conditioning of medium was started 4 days after removal of dox. For short-term p21-OE overexpression experiments shown in fig. S19, medium was allowed to be conditioned for 12 hours. For CM from *shCxcl14* knockdown cells, cycling cells were first infected with p21-OE virus for 2 days, followed by infection with *shCxcl14* virus for the next 2 consecutive days after which, on day 4, conditioning was started.

Scratch assays

Cycling P3 MEFs were seeded in 24-well plates and grown to confluence for ~3 days. Medium was removed and CM was added. Immediately afterwards, using a P20 pipette tip a linear vertical scratch was made from the top well center to the bottom well center. Cells were promptly imaged to document the initial scratch width (0 hours). Cells were grown in regular 3% O_2 incubators until 2, 12, 24, and 48 hours post-scratch when cells were imaged again. To count cells emigrating from the cell dense area into the scratch space, three to six fields of a 10X objective were quantified and invading cell number was normalized to scratch length which these cells occupied. The average scratch width was measured from two microscopy fields of a 4X objective. Per field we collected at least 10 horizontal measurements (spaced 200 μm apart) from scratch edge to scratch edge.

Isolation and characterization of peritoneal immune cells

Two-to-four-month-old wildtype mice were used to collect the peritoneal lavage (*72*) using 10 ml of ice-cold PBS applied via a 20G needle. The lavage was centrifuged at 500*g* for 10 min at 4°C. Cells were counted and subjected to transwell migration assays or used for flow cytometry. Peritoneal immune cells from wildtype control mice or wildtype mice injected with CM were resuspended in 300 μl of DMEM. A 100-μl cell suspension was used for antibody staining using eFluor 450-conjugated anti-CD11b (eBioscience, #48-0112; 1:100), FITC-conjugated

anti-B220/CD45R (BD BioSciences; #553088; 1:100), and APC-conjugated anti-TCRβ (BD BioSciences; #553174; 1:100). Cells were stained 20 min on ice in the dark, after which 200 μl of DMEM was added and cells were analyzed via a FACSCanto X (BD BioSciences). Cell counts within 60 s were noted.

Transwell migration assay

To perform transwell migration assays using peritoneal immune cells, 500 μl of CM was added to a 24-well plate. A transwell inset (3-μm pore size, Costar, #3415 or #3472) was loaded with ~2×10^5 peritoneal immune cells in 100 μl of medium (matching the medium used for CM production). Cells were allowed to migrate for 12 hours. Then, the transwell was carefully removed and the medium containing suspension cells was collected. Attached cells on the well bottom were washed twice with PBS, trypsinized and scraped. Suspension cells and attached cells were spun at 500*g* for 10 min, resuspended and counted. Cell counts were normalized to cell numbers of control condition (CM cycling cells or CM EV) for each mouse separately. For CXCL14 neutralization experiments, CM from EV- or p21-OE cells was added to a 24-well plate together with 20 μg/ml of goat anti-CXCL14 (R&D Systems, #AF866) or 20 μg/ml of goat anti-IgG (R&D Systems, #AB-108-C) (*73*). Transwell migration assays were performed as described above.

Injection of CM in wildtype mice

To determine the immune cell-eliciting potential of CM, CM was generated as described above except that culture medium with 0.5% FBS was used. One milliliter of CM was aspirated with a 25G needle and 3-ml syringe. The needle was switched to 27G and CM was slowly injected into the peritoneum of 8-10-week-old C57BL/6 wildtype mice. Four days post-injection, the peritoneal lavage was harvested and subjected to antibody staining and flow cytometry as described above.

Statistical analysis

Prism software (GraphPad Software) was used for statistical analyses. Unless otherwise stated, Student's two-tailed paired *t* tests (in MEFs and HDFs) or Student's two-tailed unpaired *t* tests (in IMR-90 cells and HUVECs) were used for pairwise significance involving two groups. For all experiments involving three or more groups, one-way analysis of variance (ANOVA) with Sidak's correction or two-way ANOVA with Sidak's or Bonferroni correction for multiple comparisons were performed. In these comparisons, the following denotes significance in all figures: $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$.

REFERENCES AND NOTES

1. L. Galluzzi, T. Yamazaki, G. Kroemer, Linking cellular stress responses to systemic homeostasis. *Nat. Rev. Mol. Cell Biol.* **19**, 731–745 (2018). doi: 10.1038/s41580-018-0068-0; pmid: 30305710
2. E. Koren, Y. Fuchs, Modes of Regulated Cell Death in Cancer. *Cancer Discov.* **11**, 245–265 (2021). doi: 10.1158/2159-8290.CD-20-0789; pmid: 33462123
3. J. M. van Deursen, The role of senescent cells in ageing. *Nature* **509**, 439–446 (2014). doi: 10.1038/nature13193; pmid: 24848057
4. T. W. Kang *et al*., Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature* **479**, 547–551 (2011). doi: 10.1038/nature10599; pmid: 22080947
5. T. Eggert *et al*., Distinct Functions of Senescence-Associated Immune Responses in Liver Tumor Surveillance and Tumor Progression. *Cancer Cell* **30**, 533–547 (2016). doi: 10.1016/j.ccell.2016.09.003; pmid: 27728804
6. T. Kuilman *et al*., Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. *Cell* **133**, 1019–1031 (2008). doi: 10.1016/j.cell.2008.03.039; pmid: 18555778
7. N. Tasdemir *et al*., BRD4 Connects Enhancer Remodeling to Senescence Immune Surveillance. *Cancer Discov.* **6**, 612–629 (2016). doi: 10.1158/2159-8290.CD-16-0217; pmid: 27099234
8. D. Muñoz-Espín *et al*., Programmed cell senescence during mammalian embryonic development. *Cell* **155**, 1104–1118 (2013). doi: 10.1016/j.cell.2013.10.019; pmid: 24238962
9. M. Storer *et al*., Senescence is a developmental mechanism that contributes to embryonic growth and patterning. *Cell* **155**, 1119–1130 (2013). doi: 10.1016/j.cell.2013.10.041; pmid: 24238961
10. A. Chiche *et al*., Injury-Induced Senescence Enables In Vivo Reprogramming in Skeletal Muscle. *Cell Stem Cell* **20**, 407–414.e4 (2017). doi: 10.1016/j.stem.2016.11.020; pmid: 28017795
11. D. J. Baker *et al*., Naturally occurring p16(Ink4a)-positive cells shorten healthy lifespan. *Nature* **530**, 184–189 (2016). doi: 10.1038/nature16932; pmid: 26840489
12. D. J. Baker *et al*., Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. *Nature* **479**, 232–236 (2011). doi: 10.1038/nature10600; pmid: 22048312
13. B. G. Childs *et al*., Senescent intimal foam cells are deleterious at all stages of atherosclerosis. *Science* **354**, 472–477 (2016). doi: 10.1126/science.aaf6659; pmid: 27789842
14. O. H. Jeon *et al*., Local clearance of senescent cells attenuates the development of post-traumatic osteoarthritis and creates a pro-regenerative environment. *Nat. Med.* **23**, 775–781 (2017). doi: 10.1038/nm.4324; pmid: 28436958
15. T. J. Bussian *et al*., Clearance of senescent glial cells prevents tau-dependent pathology and cognitive decline. *Nature* **562**, 578–582 (2018). doi: 10.1038/s41586-018-0543-y; pmid: 30232451
16. V. Krizhanovsky *et al*., Senescence of activated stellate cells limits liver fibrosis. *Cell* **134**, 657–667 (2008). doi: 10.1016/j.cell.2008.06.049; pmid: 18724938
17. J. P. Coppé *et al*., Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. *PLOS Biol.* **6**, 2853–2868 (2008). doi: 10.1371/journal.pbio.0060301; pmid: 19053174
18. A. Georgilis *et al*., PTBP1-Mediated Alternative Splicing Regulates the Inflammatory Secretome and the Pro-tumorigenic Effects of Senescent Cells. *Cancer Cell* **34**, 85–102.e9 (2018). doi: 10.1016/j.ccell.2018.06.007; pmid: 29990503
19. S. Omori *et al*., Generation of a p16 Reporter Mouse and Its Use to Characterize and Target $p16^{high}$ Cells In Vivo. *Cell Metab.* **32**, 814–828.e6 (2020). doi: 10.1016/j.cmet.2020.09.006; pmid: 32949498
20. M. Hoare *et al*., NOTCH1 mediates a switch between two distinct secretomes during senescence. *Nat. Cell Biol.* **18**, 979–992 (2016). doi: 10.1038/ncb3397; pmid: 27525720
21. J. Lovén *et al*., Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013). doi: 10.1016/j.cell.2013.03.036; pmid: 23582323
22. D. Hnisz *et al*., Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013). doi: 10.1016/j.cell.2013.09.053; pmid: 24119843
23. W. A. Whyte *et al*., Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013). doi: 10.1016/j.cell.2013.03.035; pmid: 23582322
24. J. M. Dowen *et al*., Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014). doi: 10.1016/j.cell.2014.09.030; pmid: 25303531
25. A. Chicas *et al*., Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–387 (2010). doi: 10.1016/j.ccr.2010.01.023; pmid: 20385362
26. N. Nara *et al*., Disruption of CXC motif chemokine ligand-14 in mice ameliorates obesity-induced insulin resistance. *J. Biol. Chem.* **282**, 30794–30803 (2007). doi: 10.1074/jbc.M700412200; pmid: 17724031
27. T. Hara, K. Tanegashima, Pleiotropic functions of the CXC-type chemokine CXCL14 in mammals. *J. Biochem.* **151**, 469–476 (2012). doi: 10.1093/jb/mvs030; pmid: 22437940
28. I. Kurth *et al*., Monocyte selectivity and tissue localization suggests a role for breast and kidney-expressed chemokine (BRAK) in macrophage development. *J. Exp. Med.* **194**, 855–862 (2001). doi: 10.1084/jem.194.6.855; pmid: 11561000
29. M. L. Santiago-Raber *et al*., Role of cyclin kinase inhibitor p21 in systemic autoimmunity. *J. Immunol.* **167**, 4067–4074 (2001). doi: 10.4049/jimmunol.167.7.4067; pmid: 11564828
30. B. R. Lawson *et al*., Deficiency of the cyclin kinase inhibitor p21 (WAF-1/CIP-1) promotes apoptosis of activated/memory T cells and inhibits spontaneous systemic autoimmunity. *J. Exp. Med.* **199**, 547–557 (2004). doi: 10.1084/jem.20031685; pmid: 14970181
31. D. Balomenos *et al*., The cell cycle inhibitor p21 controls T-cell proliferation and sex-linked lupus development. *Nat. Med.* **6**, 171–176 (2000). doi: 10.1038/72272; pmid: 10655105
32. L. Zitvogel, A. Tesniere, G. Kroemer, Cancer despite immunosurveillance: Immunoselection and immunosubversion. *Nat. Rev. Immunol.* **6**, 715–727 (2006). doi: 10.1038/nri1936; pmid: 16977338
33. C. Muñoz-Fontela, A. Mandinova, S. A. Aaronson, S. W. Lee, Emerging roles of p53 and other tumour-suppressor genes in immune regulation. *Nat. Rev. Immunol.* **16**, 741–750 (2016). doi: 10.1038/nri.2016.99; pmid: 27667712
34. I. Sanidas *et al*., A Code of Mono-phosphorylation Modulates the Function of RB. *Mol. Cell* **73**, 985–1000.e6 (2019). doi: 10.1016/j.molcel.2019.01.004; pmid: 30711375
35. H. Braumüller *et al*., T-helper-1-cell cytokines drive cancer into senescence. *Nature* **494**, 361–365 (2013). doi: 10.1038/nature11824; pmid: 23376950
36. S. J. Kim *et al*., Macrophages are the primary effector cells in IL-7-induced arthritis. *Cell. Mol. Immunol.* **17**, 728–740 (2020). doi: 10.1038/s41423-019-0235-z; pmid: 31197255
37. S. Boulakirba *et al*., IL-34 and CSF-1 display an equivalent macrophage differentiation ability but a different polarization potential. *Sci. Rep.* **8**, 256 (2018). doi: 10.1038/s41598-017-18433-4; pmid: 29321503
38. J. A. Westrich, D. W. Vermeer, P. L. Colbert, W. C. Spanos, D. Pyeon, The multifarious roles of the chemokine CXCL14 in cancer progression and immune responses. *Mol. Carcinog.* **59**, 794–806 (2020). doi: 10.1002/mc.23188; pmid: 32212206
39. H. J. Nam, J. M. van Deursen, Cyclin B2 and p53 control proper timing of centrosome separation. *Nat. Cell Biol.* **16**, 538–549 (2014). pmid: 24776885
40. X. Yao *et al*., Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Cell Res.* **27**, 801–814 (2017). doi: 10.1038/cr.2017.76; pmid: 28524166
41. J. R. Babu *et al*., Rae1 is an essential mitotic checkpoint regulator that cooperates with Bub3 to prevent chromosome missegregation. *J. Cell Biol.* **160**, 341–353 (2003). doi: 10.1083/jcb.200211048; pmid: 12551952
42. A. C. Flor, A. P. Doshi, S. J. Kron, Modulation of therapy-induced senescence by reactive lipid aldehydes. *Cell Death Discov.* **2**, 16045 (2016). doi: 10.1038/cddiscovery.2016.45; pmid: 27453792
43. M. Hamada *et al*., Ran-dependent docking of importin-beta to RanBP2/Nup358 filaments is essential for protein import and cell viability. *J. Cell Biol.* **194**, 597–612 (2011). doi: 10.1083/jcb.201102018; pmid: 21859863
44. J. Zhong *et al*., Purification of nanogram-range immunoprecipitated DNA in ChIP-seq application. *BMC Genomics* **18**, 985 (2017). doi: 10.1186/s12864-017-4371-5; pmid: 29268714
45. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). doi: 10.1186/gb-2009-10-3-r25; pmid: 19261174
46. Y. Zhang *et al*., Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). doi: 10.1186/gb-2008-9-9-r137; pmid: 18798982
47. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281

48. F. Ramírez *et al*., deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016). doi: 10.1093/nar/gkw257; pmid: 27079975
49. J. T. Robinson *et al*., Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011). doi: 10.1038/nbt.1754; pmid: 21221095
50. T. L. Bailey *et al*., MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009). doi: 10.1093/nar/gkp335; pmid: 19458158
51. V. Kulakovskiy *et al*., HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018). doi: 10.1093/nar/gkx1106; pmid: 29140464
52. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011). doi: 10.1093/bioinformatics/btr064; pmid: 21330290
53. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: 10.1038/nmeth.1923; pmid: 22388286
54. D. Kim *et al*., TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013). doi: 10.1186/gb-2013-14-4-r36; pmid: 23618408
55. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108–e108 (2013). doi: 10.1093/nar/gkt214; pmid: 23558742
56. A. Subramanian *et al*., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517
57. D. Merico, R. Isserlin, O. Stueker, A. Emili, G. D. Bader, Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLOS ONE* **5**, e13984 (2010). doi: 10.1371/journal.pone.0013984; pmid: 21085593
58. D. Szklarczyk *et al*., STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47** (D1), D607–D613 (2019). doi: 10.1093/nar/gky1131; pmid: 30476243
59. A. Liberzon *et al*., Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011). doi: 10.1093/bioinformatics/btr260; pmid: 21546393
60. M. Ashburner *et al*., Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000). doi: 10.1038/75556; pmid: 10802651
61. The Gene Ontology Consortium, Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017). doi: 10.1093/nar/gkw1108; pmid: 27899567
62. D. Binns *et al*., QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009). doi: 10.1093/bioinformatics/btp536; pmid: 19744993
63. A. Kahraman *et al*., TRAIL mediates liver injury by the innate immune system in the bile duct-ligated mouse. *Hepatology* **47**, 1317–1330 (2008). doi: 10.1002/hep.22136; pmid: 18220275
64. K. Aziz *et al*., Ccne1 Overexpression Causes Chromosome Instability in Liver Cells and Liver Tumor Development in Mice. *Gastroenterology* **157**, 210–226.e12 (2019). doi: 10.1053/j.gastro.2019.03.016; pmid: 30878468
65. C. N. Mayhew *et al*., Liver-specific pRB loss results in ectopic cell cycle entry and aberrant ploidy. *Cancer Res.* **65**, 4568–4577 (2005). doi: 10.1158/0008-5472.CAN-04-4221; pmid: 15930274
66. J. Moffat *et al*., A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006). doi: 10.1016/j.cell.2006.01.040; pmid: 16564017
67. T. Miest, D. Saenz, A. Meehan, M. Llano, E. M. Poeschla, Intensive RNAi with lentiviral vectors in mammalian cells. *Methods* **47**, 298–303 (2009). doi: 10.1016/j.ymeth.2008.11.001; pmid: 19041944
68. R. L. Weaver *et al*., BubR1 alterations that reinforce mitotic surveillance act against aneuploidy and cancer. *eLife* **5**, e16620 (2016). doi: 10.7554/eLife.16620; pmid: 27528194
69. M. J. Wang *et al*., Reversal of hepatocyte senescence after continuous in vivo cell proliferation. *Hepatology* **60**, 349–361 (2014). doi: 10.1002/hep.27094; pmid: 24711261
70. X. Chen *et al*., Endogenous expression of Hras(G12V) induces developmental defects and neoplasms with copy number imbalances of the oncogene. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7979–7984 (2009). doi: 10.1073/pnas.0900343106; pmid: 19416908
71. L. H. Kasper *et al*., CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol. Cell. Biol.* **19**, 764–776 (1999). doi: 10.1128/MCB.19.1.764; pmid: 9858599
72. A. Ray, B. N. Dittel, Isolation of mouse peritoneal cavity cells. *J. Vis. Exp.* (35): 1488 (2010). doi: 10.3791/1488; pmid: 20110936
73. J. Lu *et al*., IRX1 hypomethylation promotes osteosarcoma metastasis via induction of CXCL14/NF-κB signaling. *J. Clin. Invest.* **125**, 1839–1856 (2015). doi: 10.1172/JCI78437; pmid: 25822025

ACKNOWLEDGMENTS

We thank B. G. Childs, J. F. Limzerwala, C. J. Sieben, R. Bram, J. Elisseeff, and B. van de Sluis for helpful and stimulating discussions and/or careful evaluation of the manuscript. We also thank R. Velasco Fierro for valuable technical assistance. We are grateful to C. Ross for his work on the identification of super-enhancers and the genes they control, M. H. Hofker for conceptual discussions, and R. Thaler for assistance in generating heatmaps. Flow cytometry was performed by the Mayo Clinic Microscopy and Cell Analysis Core, sequencing by Mayo Clinic Medical Genomics Facility Sequencing Core, and ChIP in the Epigenomics Core of the Mayo Clinic Center for Cell Signaling in Gastroenterology. **Funding:** This work was supported by grants from Mayo Clinic's Center for Biomedical Discovery, the Paul F. Glenn Foundation for Medical Research, the Keck Foundation, the US National Institutes of Health (grants R01 AG057493, P30 DK084567 and R01 AG056318), the Mayo Clinic Cancer Center, and the David F. and Margaret T. Grohne Cancer Immunology and Immunotherapy Program. **Author contributions:** J.M.v.D. conceived the project. I.S. and J.M.v.D. designed most experiments. I.S. performed most experiments. J.-H.L. and T.O. performed ChIP-related experiments, and C.Z., I.S., and H.L. conducted super-enhancer–related studies. C.Z. and I.S. performed bioinformatic analyses. I.S., C.C.S., I.C., E.J.v.D., and J.T.S. conducted knockdown experiments, and K.B.J. coimmunoprecipitation experiments. N.H., J.G., M.H., and D.Y.L. generated and validated transgenic strains. R.M.L. helped with experimental design and data interpretation. D.F. and V.S. designed and executed neutralizing antibody experiments in mice with I.S. and helped with interpretation of immunosurveillance data. All authors contributed to data acquisition, analysis and interpretation. J.M.v.D. and I.S. wrote the paper and all authors edited the manuscript. D.J.B. helped supervise and interpret experiments pertaining to the physiological relevance of the PASP. H.L. directed, supervised, and helped interpret all bioinformatics analyses and J.M.v.D. directed and supervised all other aspects of the study. **Competing interests:** J.M.v.D. is a cofounder of Unity Biotechnology. J.M.v.D., D.J.B., and R.M.L. are coinventors on patents licensed to or filed by Unity Biotechnology. I.S., H.L., and J.M.v.D. are listed as inventors on provisional patent application U.S. 63/224,177 filed by Mayo Clinic that encompasses aspects of this publication. J.M.v.D., H.L., R.M.L., and D.J.B. are current Unity Biotechnology shareholders. This research has been reviewed by the Mayo Clinic Conflict of Interest Review Board and is being conducted in compliance with Mayo Clinic conflict of interest policies. **Data and materials availability:** ChIP-seq and RNA-seq data sets have been deposited in the Gene Expression Omnibus (GSE117278). Mouse strains *L-p21*, *L-KRAS*^{G12V}, *L-p16* and *p21*^{floxed} mice are available from Mayo Clinic upon completion of a material transfer agreement.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abb3420
Figs. S1 to S20
Tables S1 to S9
MDAR Reproducibility Checklist

23 February 2020; resubmitted 7 April 2021
Accepted 3 September 2021
10.1126/science.abb3420



AAAS.ORG/COMMUNITY







AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

RESEARCH ARTICLE SUMMARY

INDIGENOUS PEOPLES

Effects of land dispossession and forced migration on Indigenous peoples in North America

Justin Farrell*, Paul Berne Burow†, Kathryn McConnell†, Jude Bayham, Kyle Whyte, Gal Koss

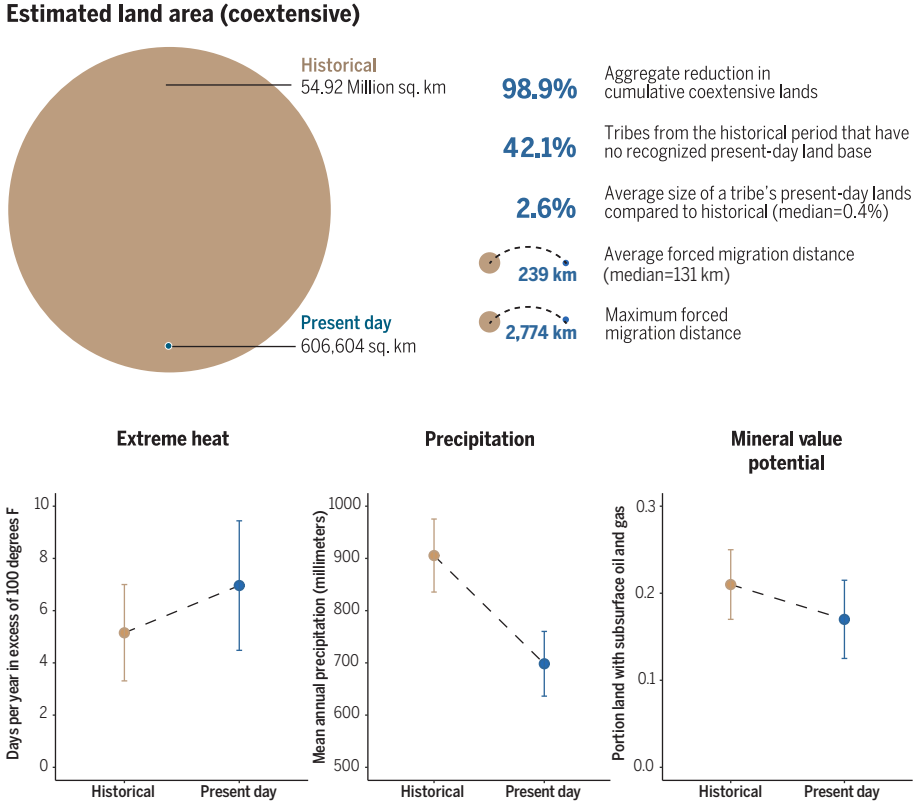
INTRODUCTION: Centuries of land dispossession and forced migration of Indigenous peoples by European and American settlers reshaped the entire North American continent. Yet the full scope of change is not quantified or systematically georeferenced at scale because of severe data constraints. Thus, fundamental questions and hypotheses still remain untested, especially concerning estimated total land loss, land value potential, and current and future climate risks. Building on historical research and Indigenous Studies scholarship, we developed a new dataset to catalog and examine the totality of land dispossession and forced migration in what is currently called the United States and tested hypotheses related to the environmental and economic impacts of these processes over time.

RATIONALE: We constructed a new comprehensive dataset compiled from a broad suite of historical sources for the vast majority of Indigenous peoples, by nation, within the boundaries of the contiguous United States. We classified the land base data for each tribe within two time periods: historical and present day. We then applied statistical models to assess two research questions at scale. First, what was the full extent of land dispossession and forced migration for each tribe and for all tribes combined? Second, did tribes' new lands, being severely reduced in size and potentially far from their ancestral lands, offer improved or reduced environmental conditions and economic opportunities over time? We tested the latter along four hypothesized dimensions that include exposure to climate change risks and hazards; mineral value potential; suitability for agriculture; and proximity to US federally managed lands that limit Indigenous movements, management, and traditional uses.

RESULTS: Statistical analysis shows that aggregate land reduction was near total, with a 98.9% reduction in cumulative coextensive lands and a 93.9% reduction in noncoextensive lands. Further, 42.1% of tribes from the historical period have no federally- or state-recognized present-day tribal land base. Of the tribes that still have a land base, their present-day lands are an average of 2.6% the size of their estimated historical area. Additionally, many tribes were forced onto new lands shared by multiple Indigenous peoples, even in cases in which nations are culturally dissimilar and have separate ancestral areas. Many present-day lands are far from historical lands. Migration dyad analysis shows that forced migration distances averaged 239 km, with a median of 131 km and a maximum of 2774 km.

Tests related to climate change risk exposure, land conditions, and potential economic value reveal substantial differences between tribes' historical and present-day areas. First, tribes' present-day lands are on average more exposed to climate change risks and hazards, including more extreme heat and less precipitation. Nearly half of tribes experienced heightened wildfire hazard exposure. Second, tribes' present-day lands have less positive economic mineral value, being less likely to lie over valuable subsurface oil and gas resources. Agricultural suitability results were mixed. Last, about half of tribes saw an increase in their proximity to federal lands in the present day.

CONCLUSION: This research suggests that near-total land reduction and forced migration lead to contemporary conditions in which tribal lands experience increased exposure to climate change risks and hazards and diminished economic value. The significance of these climate and economic effects reflect aggregate changes across the continent, but there is an urgent need to understand the magnitude of place-specific impacts for particular Native nations resulting from settler colonialism in future research. This study and dataset initiate a new macroscopic research agenda that prioritizes ongoing data collection, Tribal input, historical validation, public data dashboards, and computational analysis to better understand the long-term dynamics of land dispossession and forced migration across scales. ■



Land dispossession and forced migration impacts between historical and present-day periods. (Top left) Proportion reduction for coextensive land area estimation (accounting for multiple tribes in a single area). Limitations in the historical record likely result in an underestimation of total historical land area. (Bottom) Plots show changes in tribal land conditions (mean and 95% confidence interval) for selected variables.

The list of author affiliations is in the full article online.
*Corresponding author. Email: justin.farrell@yale.edu
†These authors contributed equally to this work.
Cite this article as J. Farrell *et al.*, *Science* **374**, eabe4943 (2021). DOI: 10.1126/science.abe4943

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abe4943

RESEARCH ARTICLE

INDIGENOUS PEOPLES

Effects of land dispossession and forced migration on Indigenous peoples in North America

Justin Farrell[1]*, Paul Berne Burow[1,2]†, Kathryn McConnell[1]†, Jude Bayham[3], Kyle Whyte[4], Gal Koss[3]

What are the full extent and long-term effects of land dispossession and forced migration for Indigenous peoples in North America? We leveraged a new dataset of Indigenous land dispossession and forced migration to statistically compare features of historical tribal lands to present-day tribal lands at the aggregate and individual tribe level. Results show a near-total aggregate reduction of Indigenous land density and spread. Indigenous peoples were forced to lands that are more exposed to climate change risks and hazards and are less likely to lie over valuable subsurface oil and gas resources. Agricultural suitability and federal land proximity results—which affect Indigenous movements, management, and traditional uses—are mixed. These findings have substantial policy implications related to heightened climate vulnerability, extensive land reduction, and diminished land value.

Historical research shows that land dispossession and forced migration are the primary means by which settler populations achieve large-scale political and economic control over Indigenous populations. Early records provide evidence of these practices as early as the 12th century BCE; followed by the Greek, Persian, and Roman empires; and more recently by European colonial expansion across the globe throughout the middle and latter period of the 2nd millennium CE. Native peoples living in what is today North America experienced large-scale forced migration and ancestral land dispossession, beginning with the arrival of European settlers and culminating in 19th- and 20th-century continental expansion by the United States.

Prior research has argued that historical land dispossession and forced migration are part of the broader sociopolitical formation of settler colonialism, which references the process through which Indigenous polities are replaced by a society of settlers whose claims to territory and governance are enabled by the extirpation of Indigenous peoples even as these communities and nations endure in the present (*1*). The primary terms we use to describe First Peoples of the continent are "Indigenous" or "Native" peoples, but we also use "tribe" or "tribal nation" throughout the text to reference both Native peoples and their polities now and in the past in the area that is currently called the United States. At times, including in our title, we reference "North America" rather than the "United States" because our data focus on sovereign tribal nations whose existence precedes the creation of the United States and its borders and still today maintains a government-to-government relationship with the US federal government. The term "North America" recognizes these nations' distinct sovereign political status even as we bound the area of study to the exterior borders of the contiguous United States. Further, many tribal nations maintain land claims that exceed the boundaries of the contemporary United States on the basis of ancestral occupancy, but we do not include those areas in order to control for the varied political-historical contexts of land dispossession in neighboring settler nation-states such as Canada and Mexico and instead focus on US policies toward Indigenous peoples within its nation-state boundaries.

Research shows that land dispossession and forced migration created the groundwork for contemporary conditions in which Indigenous peoples in the United States today face greater vulnerabilities to their health and food security, lack access to culturally appropriate education, and have heightened exposures to contaminants (*2*–*8*). Despite substantial scholarly attention to these issues, research has been primarily qualitative and based on case studies, circumscribed to detailed historical and anthropological accounts, often of an individual tribe or region (*9*–*13*). Or studies have primarily focused on rough estimates of the loss of property as measured in acres, such as the oft-cited figure that 90 million acres of land were lost through allotment of Indigenous lands under the Dawes Severalty Act of 1887 (*14*). Researchers have been unable to quantify these wide-ranging historical developments at large scales because of severe and ongoing data constraints, in which reliable information—which itself was at times suppressed during and after treaty negotiations, land ceding, and dispossession processes (*15*)—is scattered across bureaucratic agencies or buried in state, federal, and Indigenous government archives.

This lack of information creates substantial challenges when seeking to understand the collective impacts of US policies of land dispossession and forced migration on Indigenous peoples today. For without clarity and valid information on migration and dispossession, precise correlative or causal connections are challenging to justify at reliable degrees of precision.

Although prevailing qualitative approaches are indispensable and lay the groundwork for this study, their restricted scope leaves crucial questions unanswered about the totality of land dispossession and forced migration in US history and the aggregate consequences of these outcomes on Native peoples over time and leading up to the present. What was the full extent of forced migration for each tribe and for all tribes combined? What proportion of tribes no longer have any federally and state-recognized land base? We do not know the distribution of relocation distances or aggregate estimates of land area reduction, although analyses of land allotment and economic development show the substantial long-term impacts on wealth for Indigenous nations (*16*). Most critically, we only have limited systematic knowledge about the implications of the environmentally different lands that Indigenous peoples were forced to migrate to. Historical accounts show that settler governments tended to intentionally relocate tribes to what, at the time, were considered less economically desirable lands (*17*). But we do not know, for example, whether Native peoples were systematically forced to lands that are, currently, more or less vulnerable to the effects of climate change. Additionally, did these new lands come to offer Indigenous peoples improved or reduced natural resource–based economic opportunities (such as oil and gas minerals and agriculture), both at the time of relocation and into the 20th and 21st centuries as a multitrillion-dollar land-based settler economy was being built?

Regarding subsurface oil and gas resources, there is a compound issue here of US settlers creating barriers to Indigenous cultural, political, and economic self-determination and then preventing Indigenous peoples from enjoying equal benefits and safety from the emerging resource-based settler economy. But certainly, many Indigenous persons have criticized the terminology of "natural resources" as trivializing their kinship and spiritual connections to place and resist oil and gas industries on their lands. The questions here seek to examine the actions of certain US settlers who did interpret

[1]School of the Environment, Yale University, New Haven, CT 06511, USA. [2]Department of Anthropology, Yale University, New Haven, CT 06511, USA. [3]Agriculture and Resource Economics, Colorado State University, Fort Collins, CO 80523, USA. [4]School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48109, USA.
***Corresponding author. Email: justin.farrell@yale.edu**
†These authors contributed equally to this work.

lands according to certain perceived quantities and values of natural resources.

Centuries of land dispossession have reshaped an entire continent, and its effects continue to endure. Estimating these developments for the first time with a quantitative macroscopic approach will substantially broaden and deepen scientific understanding; provide a comprehensive public dataset to initiate a long-term computational research program on these increasingly important issues; complement the growing, primarily Indigenous-led efforts to map Native lands across North America and beyond (*18*–*20*); and improve future policy-making by uncovering fundamental patterns heretofore unexamined at their full geographic and temporal scope.

New dataset on Indigenous land loss and forced migration

To test these questions, we constructed a new comprehensive dataset of land dispossession and forced migration for the vast majority of Indigenous peoples in the area currently called the contiguous United States. We compiled data from many sources (table S18), including Indigenous nations' own publications and public archives, digitized administrative records of land cession treaties made between Native nations and the US government between 1722 and 1883 (*21*), judicial records from land disputes filed before the Indian Claims Commission between 1946 and 1978 (*22*), the Library of Congress and Department of Interior's schedule of Indian Land Cessions in the United States, 1784–1894 (digital scans of Congressional "Schedule of Indian Land Cessions" and "Schedule of Treaties and Acts of Congress Authorizing Allotments of Lands in Severalty") (*23*), Oklahoma State University Library's digitized "Indian Affairs: Laws and Treaties" seven-volume report from Charles J. Kappler in 1904 to the Senate Committee on Indian Affairs (*24*), the University of Nebraska Library's American Indian Treaties Portal (*25*), US Forest Service "Tribal Connections" geospatial data (*26*), the Bureau of Indian Affairs' data and reports "Indian Lands in the United States" and "Indian Entities Recognized and Eligible to Receive Services from the United States Bureau of Indian Affairs" (*27*), the Smithsonian Institution's series of reference volumes on Native American studies (*28*), and a dataset of digitized treaties from prior research (*29*). We also validated and cross-checked with crowdsourced data from the Indigenous-led organization Native Land Digital that collects, organizes, estimates, and publishes self-reported and secondary evidence of Native peoples' historical locations, including territory maps from Indigenous nations themselves (*30*). None of these sources are entirely comprehensive, but when aggregated and rigorously cross-checked, they together form a reliable state-of-the-art dataset.

We used a comparative-historical research design with two time periods: historical lands and present-day lands. We define historical lands as the earliest documented locations of Indigenous peoples in the historical data sources (table S18), often as lands held before the last 19th-century forced migration. Constructing the historical data involved difficult decisions concerning scale and historical lineage because contemporary tribes may be groups that identified as varied types of tribal formations during colonial periods even though historically their members may have associated to differing degrees with larger or more fluid social groups that were not themselves organized as "tribes." Indigenous peoples can constitute themselves by clans, kinship networks, or bands, in addition to more variable instances in which multiple tribes form a larger confederation or nation of associated tribes. For example, according to their own historical records (*31*, *32*)—and confirmed in the federal administrative records above—one-third of the Cayuga people were forced from New York to Kansas in 1846. Before European contact in 1450, the Cayuga people had become part of the Haudenosaunee confederation that included Cayuga, Seneca, Onondaga, Oneida, and Mohawk peoples and then later the Tuscarora people in 1722. Although the Haudenosaunee became one of the most important confederacies on the continent, each of these six tribes that make up this confederation are distinct peoples and are treated as such in our data. More examples and detail about the analytical, ethical, and historical difficulty of distinguishing tribes and confederacies are provided in the supplementary materials (materials and methods S7), which detail our conservative aggregate approach and future opportunities for improvement.

Any analysis of Indigenous territories must also account for the constancy of annual and interannual movements locally and regionally, diverse land tenure practices, and fluid geographic boundaries. These were widespread before and during our historical period of measurement, preceding the enforcement of predominantly fixed settler-colonial administrative boundaries in the present day. To account for these fundamental differences across time points, we tailored two different units of analysis: one for historical lands and one for present-day lands.

To measure historical lands, we used a larger unit of analysis that accounts for diverse land tenure practices; accommodates systematic tribal movements, coextensive land areas, and shifting boundaries; and avoids the problematic enforcement of overly rigid perimeters prone to measurement error through underestimation. Because of the historical predominance of settler-colonial administrative boundaries, most national-level administrative data have been continually collected at the US county and county-equivalent (CCE) level rather than the Indigenous governmental level. For the purposes of quantitative analysis, CCEs are especially useful and necessary to examine historical Indigenous territories because these administrative boundaries are often large enough to avoid such measurement error (spatial underestimation) but still defined enough to allow for spatial variation to make large-scale comparisons by using a host of reliable administrative and environmental data.

We paired the Indigenous and other historical sources above with data from the US Department of Interior and US Department of Agriculture (USDA) Forest Service—produced under the Native American Graves Protection and Repatriation Act (*33*, *34*)—that identified the present-day states and counties included wholly or partially within a map cession boundary (*35*). Using this multisourced process, we geocoded and plotted all historical Indigenous lands within CCEs, creating a binary indicator of whether or not a tribe had territory in a CCE.

Indigenous land tenure systems were varied, so these boundaries create room for multiple modes of occupancy and territorial claiming necessary to construct this dataset and to conservatively—and most reliably—make temporal comparisons about land qualities and climate. And although CCEs proved to be the most methodologically reliable option for large-scale historical study, it is essential to recognize that the restriction of Indigenous peoples to a bounded area is itself a practice and outcome of colonial forced migration (*36*, *37*). Further, because some tribes have systematic migratory land tenure practices, it is impossible and analytically unsound to impose overdetermined historical boundaries. Moreover, even attempts to identify exact historical borderlines again reinforces settler-colonial assumptions about modes of land occupancy, private-property regimes, and permanent geographic limitations not shared by some Indigenous peoples.

Although far-reaching and accurate, continued data collection from oral histories, tribal records, and archaeological records will be needed to elongate and deepen the historical reach of the data, especially for adding smaller unrecognized tribes, communities, and Indigenous homelands. Any missing tribes are a limitation of the historical record and publicly available data and do not constitute definitive evidence relevant to any conflicts or legal claims about the political recognition of particular Indigenous peoples. Only tribes themselves can provide decisive historical boundaries when and where they exist. Further, there are some cases in which the historical data cannot disentangle the forced migration routes of tribes who moved numerous times under such conditions, which can create an incomplete history of when and from where they migrated.

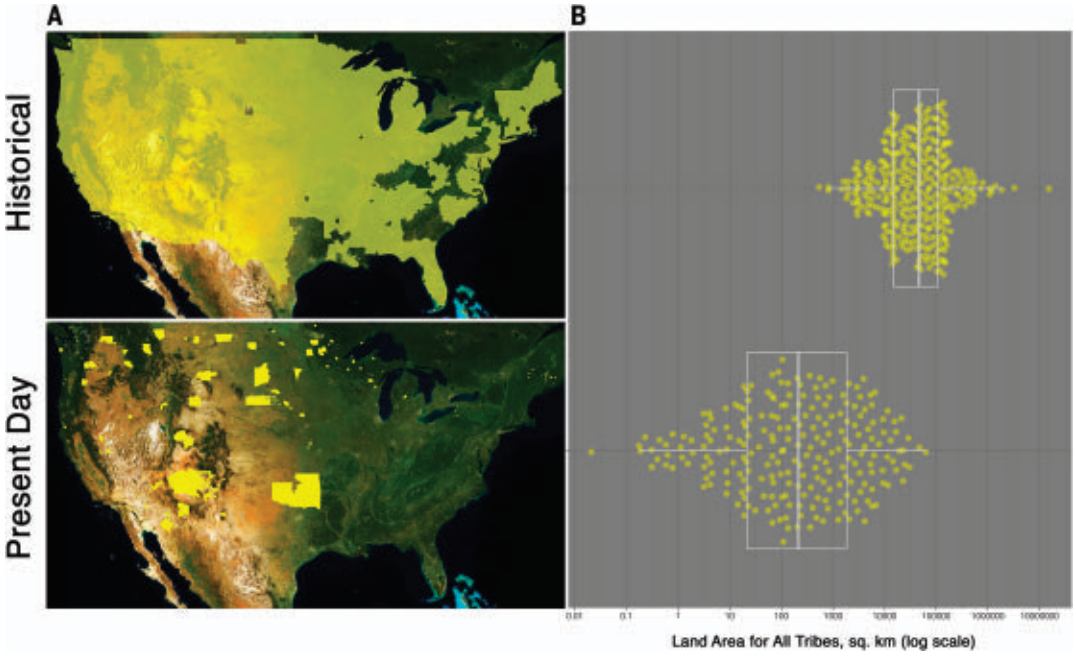


Fig. 1. Historical and present-day locations for all tribes in the data, revealing a sharp decline in land coverage. (Top) Historical. (Bottom) Present day. (**A**) Map showing the aggregate land base for all tribes. (**B**) Plot comparing the total land area of each tribe, revealing a sharp decline from historical territories to present-day lands. Land area values for tribes that drop out between the historical and present-day periods are coded as zero to capture cases of total land base loss. Unshaded areas in both images still have Indigenous historical or present-day land tenure relevance, including for future Indigenous cultural and political self-determination. And limitations in the historical record likely underestimate Indigenous presence in the historical map and can exclude identifiable references to landless or unrecognized tribes. More on this point and the expanded statistical results are provided in table S4 and materials and methods S7.

Although our current analysis is based on data compiled with these historical limitations, the data still capture a preponderance of known cases of Indigenous land dispossession.

For the present day, we measured lands using the finest-grained distinct geographic unit of analysis made available by the US Census Bureau, which is composed of tribal census blocks and block groups that can be aggregated to the tribe level. We used the well-established "American Indian/Alaska Native/Native Hawaiian (AIANNH) Area National Shapefile," which contains precise polygon records for legal and statistical entities, including all federally recognized American Indian reservations and off-reservation trust land areas as well as state-recognized American Indian reservations (*38*). These data include Oklahoma Tribal Statistical Areas (OTSAs), Tribal Designated Statistical Areas (TDSAs), and State-Designated Tribal Statistical Areas (SDTSAs). The relevance of these different land categories is discussed further in the supplementary materials, materials and methods S7.1. These census data allow us to measure present-day tribal lands with precision and reliability.

We then manually paired every tribe's historical land location with its present-day census block group data, a research-intensive process that included matching many tribal and subtribal name variations and locations across historical and present-day time periods to create a total of 616,157 records (full descriptive statistics are provided in tables S1 to S4, and more detailed data collection methods and validation processes are provided in materials and methods S7). In developing and combining these data, we have constructed the most comprehensive paired collection to date, with two tailored units of analysis that ethically and accurately allow for statistical modeling to assess forced migration at a large scale and to compare features of historical lands to present-day lands.

Quantifying land reduction and forced migration

The geographical and temporal distributions of land coverage for all tribes in the data are shown in Fig. 1 and table S4, revealing highly significant differences between historical and present-day lands. We represent aggregate historical land coverage in two different ways to account for the fact that multiple tribes can occupy one territory either at the same point in time or during different historical periods. First and most basically, we computed the total spatial land coverage (Fig. 1A) by summing the area of all polygons with at least one tribe known to be present. This does not consider multiple tribes in a single area but is a more precise estimation of geographical spread across the entire continent. In the historical period, tribes had a documented presence in 7,011,450 km^2 of the area of what is currently called the contiguous United States (Fig. 1A, top). In the present day, tribes had a formally recognized presence in 426,598 km^2 of the area of the contiguous United States—a reduction of 93.9% ($P < 0.001$). Limitations in the historical record likely underestimate Indigenous presence in the historical map (Fig. 1A, top), and unshaded areas in Fig. 1A, top and bottom, still have Indigenous historical or present land tenure and importance, including for future Indigenous cultural and political self-determination.

Second, because some areas were more densely populated with multiple tribes, we computed a cumulative sum of all known historical land areas for all tribes. A single historical location is counted multiple times consistent with the number of tribes with a documented historical presence in it. This measure better accounts for the amount of coextensive shared land across tribes and for the systematic tribal movements and shifting boundaries that characterized the historical period. Across all tribes in the historical period, there was a documented association with 54,919,152 km^2 of land—a figure that was reduced by 98.9% in the present day.

We focused primarily on the differences between known historical and present-day tribal lands, yet notably, 42.1% of all tribes in the dataset with documented historical presence had no federally- or state-recognized land base in the present day. For these tribes, the reduction in their land base was total. Further, in addition to the issue of land area reduction is the issue of the contraction in the number of tribes themselves between the historical period and the present day. Among tribes that still had a land base in the present day (58.2%), their contemporary lands were significantly reduced (Fig. 1B). On average, tribes' present-day lands are 2.6% the size of their estimated historical areas (median = 0.4%). In raw terms, on average these tribes saw a reduction of 215,308 km^2 from

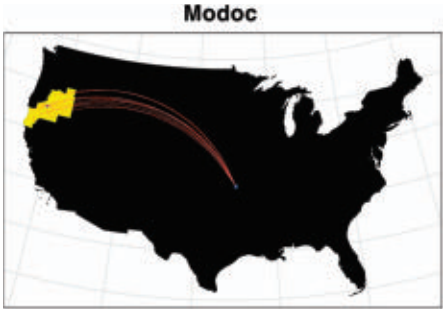


Fig. 2. Illustration of geographic dispossession and forced migration by using a sample tribe from the data, the Modoc people. These are not step-by-step migration points but are basic dyads comparing in yellow (*1*), lands documented to be held before the last forced migration, with in blue (*2*), current jurisdictional homelands in both Oregon and Oklahoma. Red arcs help to illustrate general land reduction and forced migration patterns.

the documented historical period (median = −67,289 km^2). Without exception, every present-day tribe has a smaller land base than they did in the historical period. Additionally, many tribes were forced onto lands shared by multiple Indigenous peoples, even in cases in which nations were culturally dissimilar and ancestral areas separate.

After demonstrating aggregate levels of land reduction, we next analyzed forced migration and its consequences. To do this, we created a migration dyad for every Indigenous nation, comparing all historical land areas to all present-day land areas. Because some tribes were historically spread across many areas—sometimes involving great distances—we recorded every possible migration dyad (n = 716,856) between every possible historical and present-day location for each tribe. For all possible dyads, the average distance for a tribe between historical and present-day lands was 239 km, with a median of 131 km. The Kickapoo people recorded the furthest total average distance (1366 km), and the Modoc people (Fig. 2) recorded one of the longest single instances from historical lands in the Klamath Basin of what is now California and Oregon to Ottawa, Oklahoma (2565 km). We detailed these migration dyad results with the recognition that not all tribal members of a given Native nation, either now or in the past, live exclusively on tribal lands. But as an approximation of where people of a given tribe live, changes in the location of the tribal land base still accurately demonstrate overall spatial patterns of forced migration.

Long-term impacts: Climate change exposure and natural resource endowments

Did the new lands, despite being severely reduced in geographical density and spatial coverage, offer tribes improved or reduced value and opportunity? We do not ask this question in the spirit of the diverse values that tribal members held at the time regarding their relationships with land. Rather, we ask this question with regard to the values of the territorial economy that US settlers and the federal government forcibly created. We assessed the degree to which tribes were advantaged or disadvantaged in the wake of land dispossession to participate equally in that economy. To investigate this, we compared each tribe's historical lands to their present-day lands across multiple environmental and natural resource dimensions.

We examined four critical dimensions: (i) exposure to climate change risks and impacts, (ii) mineral value potential of lands, (iii) suitability for agriculture, and (iv) proximity to US federally managed lands. The primary difficulty of comparing historical lands to present-day lands is that nearly all land in the United States has been affected by resource development, industrialization, and other US-sanctioned economic development activities during the 19th and 20th centuries and up until the present. We overcame these methodological difficulties by using indicators that are largely durable across time or not heavily affected by the human inhabitants (annual precipitation, oil and gas resources contained within hydrocarbon fields and sedimentary basins below the Earth's surface that formed over hundreds of millions of years, elevation, and terrain ruggedness). However, we purposefully include proximity to US federal lands because that is mutable across time, allowing us to test hypotheses concerning the potential social and political consequences of present-day tribal lands' proximity to these protected lands. Last, we precisely measured these four dimensions using reliable indicators at fine spatial resolution, which we aggregated at the CCE and tribal census block level.

First, we examined climate change risks and impacts using four indicators: extreme heat, long-term drought severity, annual precipitation, and wildfire hazard potential. Our measure for extreme heat is the average of days per year with a maximum temperature over 100°F (~38.8°C)—a threshold that research shows to have especially deleterious health impacts (*39*)—constructed by using data from gridded (4-km resolution) gridMET (*40*, *41*). We measured long-term drought severity using weekly Palmer Drought Severity Index (PDSI) data from *gridMET* (*42*) that includes mean and median weekly drought conditions and "captures the basic effect of global warming on drought through changes in potential evapotranspiration" (*43*). Annual precipitation is also affected by climate change (*43*) and is reliably correlated with agricultural value in nonirrigated arid areas (*44*), and thus we include a standard measurement of annual precipitation based on the standard 30-year normals (1981–2010) developed by the PRISM Climate Group at Oregon State University (*45*). Last, climate change is also linked to increasing wildfires in the United States (*46*), affecting ecosystem services, local economies, air quality, and human health. We constructed a measure of mean and median wildfire hazard potential for all grid cells within a CCE using the gridded (270 m) index from the US Forest Service Wildfire Hazard Potential (WHP) data (*47*).

Noticeable differences between historical and present-day lands are shown in Fig. 3 across several measures of climate risk exposure and mineral potential. The distribution of extreme heat days is shown in Fig. 3A in log scale. Although the mode of the historical and present-day land distributions is similar, tribal lands in the present day experience more extreme heat days. We formally tested whether the distributions are different using a Kolmogorov-Smironov (KS) test. The KS test indicates that the historical and present-day distributions are different with a statistic of 0.14 ($P < 0.05$). A detailed description of the methods is available in the supplementary materials. We supplemented our analysis with a series of regression models designed to test the difference in mean heat days in the historical and present-day lands (detailed descriptions and robustness checks on every model are provided in tables S5 to S17). All inference is based on robust standard errors. We implemented several alternative specifications and included the Bonferroni adjustment to correct for testing multiple hypotheses. Our results indicate that present-day lands endure nearly two additional extreme heat days per year compared with their historical lands [1.81, 95% confidence interval (CI): 0.49, 3.13].

In addition to extreme heat, we evaluated other measures of climate change risk exposure, including drought, precipitation, and wildfire hazard potential. We found that the present-day distributions are statistically different from the historical distributions for drought (KS statistic 0.55) and precipitation (KS statistic 0.22). Average annual precipitation has declined by nearly 23% in the present-day lands relative to historical lands (−207 mm, 95% CI: −270.00, −144.48) (Fig. 3B); 36.3% of tribes experienced increased drought between the historical and present-day. Aggregate mean drought conditions decreased (0.40 PDSI, 95% CI: 0.26, 0.53). Last, we did not find aggregate mean differences for wildfire hazard potential (0.03, 95% CI: −0.07, 0.14). Although the historical and present-day wildfire hazard potential distributions are not statistically different, the longer right-tail distribution of greater wildfire hazard in the present day demonstrates that many tribes do face significant wildfire risk today. And our data show that

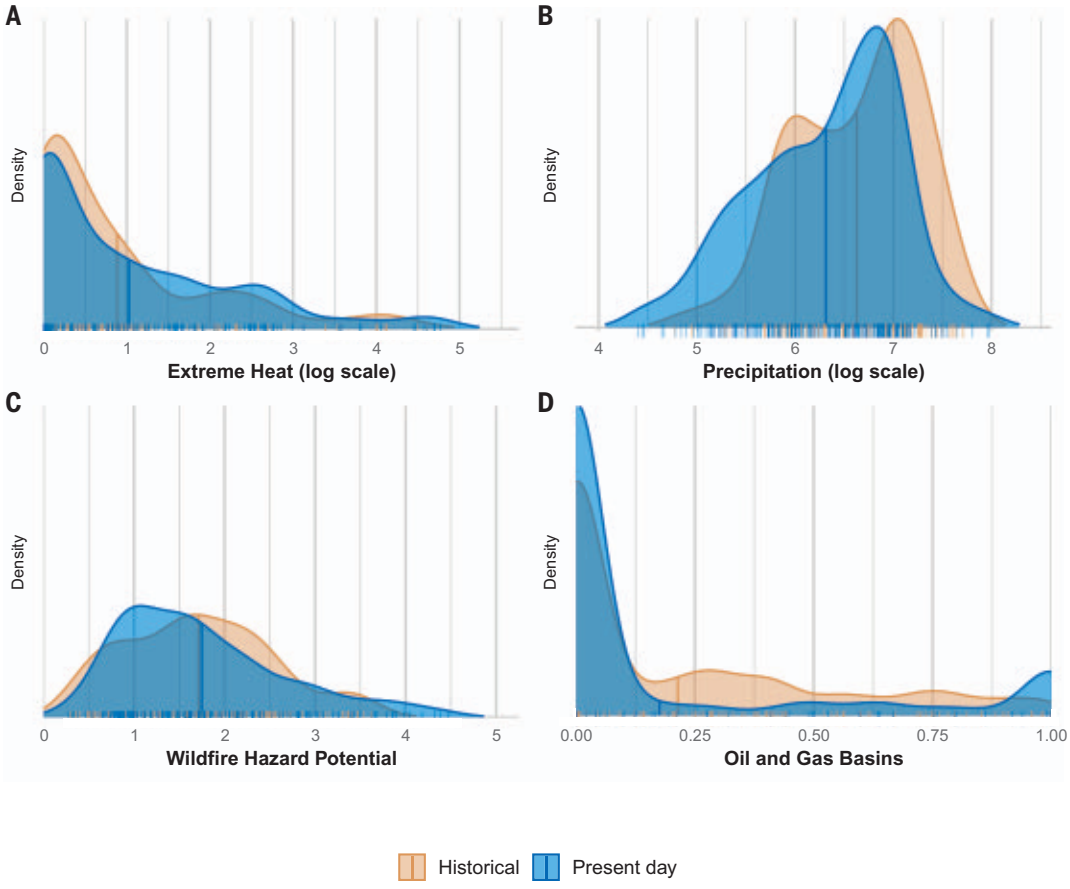


Fig. 3. Plots of differences between tribes' historical and present-day lands. (**A** to **D**) Descriptive increases on present-day lands. (A) Days of extreme heat (temperature above 100°F/~38.8°C), in log scale. (B) Precipitation (millimeters), in log scale. (C) Wildfire hazard potential (scale from very low to very high, with 5 being highest risk). (D) Decreases in oil and gas basins (proportion of tribal land on top of oil and/or gas basin). Means are indicated in all plots with vertical lines. Plots for all variables are available in fig. S1.

47.7% of tribes saw increased wildfire risk on present-day lands relative to their historical lands. Given that wildfire hazard reflects both climatic and anthropogenic factors tied to settler land management regimes, these results offer initial insight into the acute vulnerability that many tribal nations face from the risk of wildfire.

Although present-day lands endure more severe heat and less precipitation, we also sought to assess the positive economic mineral value potential beneath these lands. Again, positive economic mineral value is understood in terms of values tied to the territorial economy it forcibly created. Like our climate change measures, our primary indicator of mineral value is durable across time, recording the fraction of land that sits on top of a subsurface sedimentary basin containing oil and gas minerals. Data for this variable come from the US Energy Information Administration's shapefile of all sedimentary basin boundaries in the continental United States (*48*). The distribution of lands overlying sedimentary basins is lower in present-day lands than historical lands (KS statistic, 0.23) (Fig. 3D). Because of the disproportionate amount of oil and gas beneath Oklahoma Tribal Statistical Areas, we examined statistical differences with and without the OTSAs in the model (table S13, models 1 and 3). In our main model (table S13, model 1), the results indicate that present-day lands are 19% less likely to lie over subsurface oil and gas resources (−0.04, 95% CI: −0.07, −0.005), and when we exclude OTSAs (table S13, model 3), present-day lands are 24% less likely to lie over subsurface oil and gas (−0.04, 95% CI: −0.07, −0.01). These findings provide evidence that tribes were moved to lands with less mineral value potential.

To supplement this finding, we included two auxiliary variables that are highly correlated with the location of oil and gas sedimentary basins but instead are contingent on social and political conditions over time. We included these because they may provide a slightly different means to understand the historical long-term effects of forced migration on the positive mineral value of tribal lands. Although we address the presence of resource endowments, our analysis does not capture the institutional factors that are shown to limit use of resource endowments, such as land ownership, fragmented patterns of land ownership, fractionation, or complex property rights regimes common to many Indigenous nations in the United States (*48*, *49*). Our measures come from US Geological Survey data that record the presence (in cells of one-quarter square miles) and production status of every oil and gas well drilled between 1859 and 2005 (*50*). These data, which include more than 3 million wells, were overlaid on our tribal lands data to create two variables—one for the portion of lands with oil-producing wells and another for the portion of lands with gas-producing wells—which together are a reliable measure of energy development on every individual land area over 146 years.

The results indicate that the majority of present-day tribal lands have a very small proportion of land with oil-producing and gas-producing wells. The KS test suggests that the present-day distributions differ from the historical distribution (oil KS statistic: 0.40; gas KS statistic: 0.54). With the OTSAs included, the difference between the present-day and historical mean proportion of land having actively producing oil or gas wells is not statistically significant. When we excluded OTSAs, we found significant declines for oil production, declining 50% (−0.01, 95% CI: −0.02, −0.002). These supplementary findings suggest that land dispossession limited tribes' capacity to participate in the resource extraction economy to the same degree as that of settlers.

Next, we assessed the suitability of land for agriculture, an essential human activity that sustains societies and produces economic value for communities (*51*). Plant cultivation practices

vary across human history, but all require access to land. Although near-total land reduction (Fig. 1 and table S4) affects future agricultural possibilities more than any other factor, examination of such possibilities is still of value, even on severely reduced land areas. Research has shown that land with certain physical characteristics—such as its elevation, terrain ruggedness, and soil properties—plays a critical role in making agriculture possible, productive, and transportable to markets (*52*, *53*). We measured these physical factors in ways that are largely durable across time, enabling us to compare historical lands with present-day lands. We used terrain elevation data from the US Geological Survey (USGS) (*54*) to estimate changes in elevation and landscape ruggedness. Ruggedness is a common measure that describes the degree of topographic heterogeneity in an area—for example, a steep, mountainous region would be considered highly rugged. Thus, lands characterized by increased terrain ruggedness limit economic activity by inhibiting the mobility of goods and people, just as higher-elevation lands are generally less suitable for agricultural development, depressing its economic potential. The durability of soil organic carbon is questionable given the impact of land-use practices, but we cautiously include it here with that caveat (materials and methods S2.7). We found that present-day lands lie at nearly 21% lower elevation than that of historical lands (−155.65 m, 95% CI: −188.38, −122.93), and present-day lands are concentrated in 29% less rugged terrain (−2.84, 95% CI: −3.70, −1.97]). We found that soil organic carbon is similar in the present-day and historical lands (−0.18, 95% CI: −0.42, 0.07). However, we did find a 28% reduction in soil organic carbon in present-day lands when focusing only on OTSA tribes (−0.61, 95% CI: −0.88, −0.33). These results suggest that present-day lands may be, on average, more suitable for agriculture than were historical lands. However, a major caveat and alternative interpretation of these results is that Indigenous historical land areas were not only much larger but that these declines in ruggedness and elevation suggest that their historical lands spanned areas with more topographical diversity on average, such as high-mountain areas in the western half of the continent, many of which are now publicly owned and predominantly managed by the US Department of Interior and USDA Forest Service. These findings should also be interpreted alongside the findings above that present-day lands receive less precipitation and experience more extreme heat—and more fundamentally, alongside the near-total reduction in land described here, given that access to land at scale is itself necessary for agriculture.

Last, we further examined the effects of forced migration by bringing these data to bear on longstanding questions about the relationship between tribal lands and more than 640 million acres of lands that would come to be claimed and managed by the US federal government. The majority of these lands—which make up nearly 30% of all land in the United States—are handled by four agencies: the Bureau of Land Management, US Forest Service, US Fish and Wildlife Service, and National Park Service (*55*). Using the USGS Protected Areas Database and shapefile, we constructed a measure that records the total proportion of federally protected land within 160 km (100 miles) for all land areas in the tribal data (*56*). Historically, these lands that would become US "public lands" were initially viewed by settlers as empty land and of little intrinsic or economic value before the conservation movement in the late 19th century, which ushered in new institutions and legal regimes for their protection and management often predicated on the dispossession of Indigenous lands (*57*–*59*).

Building on our findings above about total land reduction (Fig. 1 and table S4), we hypothesized that forced migration patterns may have resulted in tribes being moved to these areas that were geographically rural, "unsettled," and perceived to be of less economic value. Our results suggest that the present-day distribution of lands with federally managed land within 160 km differs from the historical distribution (KS statistic 0.2). A comparison of density plots (fig. S1) reveals that present-day tribal lands are multimodally distributed, suggesting that one group of tribes is proximate to a large portion of federal lands, whereas others are not. Aggregate mean differences are nonsignificant [0.001, 95% CI (−0.01, 0.01)], but we found that 46.2% of tribes saw an increase in their proximity to federal lands in the present day, compared with historical lands.

Further work is needed to elucidate the distribution of change in particular places in the present day. Nevertheless, in line with previous historical research (*57*–*59*), it provides initial evidence that some tribes were forcibly moved to historically less desirable and less valuable areas of the continent. It also suggests that by being moved, some tribes were now more proximate to neighboring lands that would come to have rigid bureaucratic boundaries that restrict traditional tribal movements, management, and ecological uses. However, the proximity of Indigenous lands to federal lands, although frequently restricting important cultural uses, may in some cases protect areas from natural resource development depending on its land status and management, allowing for uses that are disrupted by private property ownership. Additionally, federal lands may offer the opportunity for tribal land recovery on a larger scale than the purchase of private lands at fair market value. Nevertheless, dispossession and US settler land management required taking on government structures stemming from the Indian Reorganization Act, the Indian Self-Determination and Education Assistance Act, and other laws and policies (*60*–*64*). During the Indian New Deal period (1930s), and later with what is sometimes called the tribal self-determination era (1970s to present), tribes were given incentives and threatened penalties if they did not organize as a formal corporation or as a state-like bureaucratic entity. Tribal governments must operate under strict rules and regulations that some have argued are unfair. The breakup of reservations led to jurisdictional situations of high bureaucratic and managerial complexity for some tribes because their lands involve fragmented trust and fee lands (for example, checkerboarding), as well as the presence of properties with fractionated ownership (for example, multiple owners). The US management of neighboring lands was now controlled by federal "multiple use" policies intended to maximize value for the US territorial economy from which settlers sought to benefit, through timber harvest, recreation, livestock grazing, and oil and gas development (*65*).

Discussion

This large-scale quantification and georeferencing of land dispossession and forced migration provides macroscopic empirical insight into ongoing efforts across many fields to better understand the path dependencies that have led to unequal distributions of socio-environmental risks and rewards. Our results show that in addition to a significant aggregate reduction in land density and spread, Indigenous peoples were pushed to lands that are now more exposed to climate change hazards; less likely to lie over subsurface oil and gas resources; and many tribes saw an increase in proximity to federally managed lands that may limit traditional tribal movements, management, and uses. Many tribes saw their entire land base dispossessed, resulting in the complete elimination of potential environmental amenities and risk exposure that may have come with a tribal land base, in addition to the fundamental importance of land and political recognition for Indigenous self-governance, cultural practices, and social identities. Although these results are of substantial relevance to many fields across the qualitative and quantitative social sciences, we also present this study as a new methodological paradigm and agenda for continued large-scale empirical investigation into these relationships.

Because of the chronological, qualitative, and ethical difficulties of collecting and validating data at this scale, this study still faces quantitative limitations imposed by the historical record, and thus, research must prioritize

ongoing collaboration, collection, and refinement that builds continually from tribal oral histories, reliable crowdsourcing, and Indigenous and settler archival records. Until now, there have been no academic-level repositories to encourage such large-scale aggregation and collection of multiple sources of data. The new dataset is now online and includes a function by which tribal members and others with historical knowledge can submit geographical additions that will be validated, integrated, and openly shared. Reliable data updated in perpetuity from multiple sources will enable future work to improve on this initial attempt and is critical for the continued development of a new macroscopic research agenda. In the future, it will be particularly important to include First Nations and other Indigenous polities in the regions that now include Canada and Mexico because many nations have homelands that transcend contemporary settler nation-state boundaries.

The findings suggest that as society seeks to address its greatest problems—climate change, land degradation, and economic and social inequality—we ought to construct policies with the scientific knowledge that these problems disproportionately affect Indigenous peoples. As a result of land dispossession and forced migration, Indigenous peoples now face increased climate vulnerability, diminished economic value of their lands, and for some tribes, restrictive federal land management regimes, which included strategic exclusion by the United States from equal participation in the emerging energy and industrial sectors that the US federal government valued and became pivotal to its future territorial economy. Science can never quantify the depth of social and cultural trauma, economic deprivation, and political marginalization of losing entire land bases and ecosystems. But the results here contribute fundamental knowledge about the factors that produced current inequities and future risks and should be especially useful for creating policies for restitution. For example, proximity to federal lands may offer an opportunity and institutional pathway for the restoration of erstwhile tribal lands (*66*–*68*).

Our results challenge common approaches to climate change policy-making, revealing the need to better integrate ethics and justice principles by addressing the disproportionate harms that climate destabilization, federal land management, and energy industry operations cause to Indigenous peoples. National-level climate policy often takes an ahistorical approach to mitigation and adaptation activities, emphasizing net reductions in greenhouse gases. Yet, for Indigenous peoples, historical land dispossession and forced migration themselves have created the conditions that intensify climate change vulnerability and risks. These factors include confinement to lands with heightened vulnerability to climate change, restrictions on the mobility of Indigenous peoples to exercise important adaptation options, and degradation of lands because of heightened fossil fuel and other natural resource extraction activities. The immense scale of land dispossession and forced migration provided the settler land base for widespread fossil fuel extraction across North America, which in turn has generated harms and risks to tribal homeland jurisdictions (*69*). Indigenous ancestors did not consent genuinely or at all to land dispossession and forced migration. Their descendants do not consent to the idea that the legacies of historical wrongdoing should today be considered acceptable or beyond redress (*70*–*72*).

There is a remaining climate-related responsibility on the part of nation-states to acknowledge and mediate the most harmful results of dispossession and forced migration. Yet with any policy debate about reparations or restorative justice, one major issue is how to make an accurate assessment of the scope of the historical wrongdoing.

This study shows that the heightened climate vulnerability of Indigenous peoples can be attributed to prior national policies and actions that resulted in massive changes in land tenure and land use. Thus, addressing climate change impacts requires not only attention to the immediate or future impacts of discrete environmental changes but also recompense for past policies and actions that continue to burden particular groups, such as Indigenous peoples. Although decisions to dispossess land from Indigenous peoples were not made with an understanding of future rises in global average temperature, they were nonetheless intended to remove Indigenous peoples as barriers to the interests of settler and migrant populations and foreign investors.

In the most recent US National Climate Assessment (*73*), the "Tribes and Indigenous Peoples" chapter focused on documenting vulnerability to climate change. The cited literature references scientific and tribal perspectives that suggest that land loss or forced migration is at the heart of vulnerability to climate change. An implication is that climate-adaptive responses would involve addressing the particular climate-related issue, such as coastal erosion or extreme heat, but would also address factors affecting landscape resilience that are rooted in historical land dispossession and forced migration. Yet those working at federal, tribal, state, or local government or working for nongovernmental organizations and universities often do not have an accurate grasp of the details, scale, and scope of dispossession and migration. This study represents a new macro-level attempt to provide such information at a large scale and serves as a basis not only for ongoing efforts to mitigate future impacts of climate change but also for new policies to remediate the historical causes responsible for generating vulnerability in the first place.

Materials and methods summary

We constructed the dataset using a wide range of publicly available historical settler and tribal sources, a process involving substantial analytical difficulties and ethical concerns laid out in the main text above and in greater detail in the supplementary materials. Historical estimates for every tribe were linked qualitatively, case by case (616,157 records), with the more spatially resolved present-day tribal census block and block group data. Our land characteristic statistical analysis uses the tribal area and time period as the unit of analysis, but tribe- and reservation-level data are also available in the dataset. Summary statistics for all variables are presented in the supplementary materials, materials and methods S1 and S2. We made historical comparisons using a series of regression models and KS tests. We implemented several alternative specifications, and inference is based on robust standard errors (tables S5 to S17). A detailed account of all data and methods is provided in the supplementary materials.

REFERENCES AND NOTES

1. P. Wolfe, Settler Colonialism and the elimination of the native. *J. Genocide Res.* **8**, 387–409 (2006). doi: 10.1080/14623520601056240
2. P. J. Deloria *et al*., Unfolding futures: Indigenous ways of knowing for the twentieth-first century. *Daedalus* **147**, 6–16 (2018). doi: 10.1162/DAED_a_00485
3. J. K. Maldonado, C. Benedict, R. Pandya, Eds., *Climate Change and Indigenous Peoples in the United States: Impacts, Experiences and Actions* (Springer, 2014).
4. J. M. O'Brien, *Dispossession by Degrees: Indian Land and Identity in Natick, Massachusetts, 1650–1790* (Cambridge Univ. Press, 1997).
5. Indian Lands in Indian Hands, Indian Land Tenure Foundation (2020); https://iltf.org.
6. D. L. Fixico, *The Invasion of Indian Country in the Twentieth Century: American Capitalism and Tribal Natural Resources* (Univ. Press Colorado, 2012).
7. R. Thornton, *Studying Native America: Problems and Prospects* (Univ. Wisconsin Press, 1998).
8. V. Deloria Jr., *Custer Died for Your Sins: An Indian Manifesto* (Univ. Oklahoma Press, 1988).
9. S. Banner, *How the Indians Lost Their Land: Law and Power on the Frontier* (Harvard Univ. Press, 2007).doi: 10.4159/9780674020535
10. D. Chang, Enclosures of land and sovereignty: The allotment of American Indian lands. *Radic. Hist. Rev.* **109**, 108–119 (2011). doi: 10.1215/01636545-2010-018
11. S. L. Egbert *et al*., Territorial cleansing: A geopolitical approach to understanding mass violence. *Territ. Politic. Gov.* **4**, 297–318 (2016). doi: 10.1080/21622671.2015.1036911
12. E. Greenwald, Reconfiguring the reservation: The Nez Perces, Jicarilla Apaches, and the Dawes Act (Univ. New Mexico Press, 2002).
13. K. Ruppel, *Unearthing Indian Land: Living with the Legacies of Allotment* (Univ. Arizona Press, 2008).
14. J. McDonnell, *The Dispossession of the American Indian, 1887–1934* (Indiana Univ. Press, 1991).
15. C. Ryan, The written record and the American Indian: The archives of the United States. *West. Hist. Q.* **6**, 163–173 (1975). doi: 10.2307/967979
16. B. Leonard, D. P. Parker, T. L. Anderson, Land quality, land rights, and indigenous poverty. *J. Dev. Econ.* **143**, 102435 (2020). doi: 10.1016/j.jdeveco.2019.102435
17. C. Saunt, *Unworthy Republic: The Dispossession of Native Americans and the Road to Indian Territory* (W.W. Norton, 2020).

18. T. Ahtone, R. Lee, K. Goodluck, M. Pearce, Land-grab universities. *High Country News* 30 March 2020; www.hcn.org/issues/52.4/indigenous-affairs-education-land-grab-universities.
19. National Tribal Geographic Information Support Center; https://tribalgis.com.
20. Native Land Information System; www.nativeland.info.
21. US Government Printing Office, *"Indian Land Cessions in the United States," Eighteenth Annual Report of the Bureau of American Ethnology to the Secretary of the Smithsonian Institution, 1896–1897* (Government Printing Office, Washington, DC, 1895); https://library.si.edu/digital-library/book/annualreportofbu182smit.
22. National Archives, "General Records of the Indian Claims Commission," RG 279 (National Archives Building, Washington, DC).
23. The Library of Congress, *Indian Land Cessions in the United States, 1784–1894* (2020); https://memory.loc.gov/ammem/amlaw/lwss-ilc.html.
24. United States, Indian Affairs: Laws and Treaties (Government Printing Office, Washington, DC, 1903).
25. University of Nebraska-Lincoln, American Indian Treaties Portal; http://treatiesportal.unl.edu.
26. US Forest Service, Indian Lands Map Viewer. Tribal Connections; http://usfs.maps.arcgis.com/apps/webappviewer/index.html?id=fe311f69cb1d43558227d73bc34f3a32.
27. US Bureau of Indian Affairs, "Indian Entities Recognized by and Eligible To Receive Services From the United States Bureau of Indian Affairs," 84 FR 1200 (Government Printing Office, 2019); www.federalregister.gov/d/2019-00897.
28. W. Sturtevant, Ed., *Handbook of North American Indians* (US Government Printing Office, 1978), vol. 2–17.
29. A. Spirling, U.S. treaty making with American Indians: Institutional change and relative power, 1784–1911. *Am. J. Pol. Sci.* **56**, 84–97 (2012). doi: 10.1111/j.1540-5907.2011.00558.x
30. Native Land Digital; https://native-land.ca.
31. Cayuga Nation, History Timeline; https://web.archive.org/web/20200708204541/https://cayuganation-nsn.gov/history-timeline.html.
32. T. Shannon, *Iroquois Diplomacy on the Early American Frontier* (Penguin Books, New York, 2009).
33. National Park Service, US Department of the Interior, Native American Consultation Database, National NAGPRA Online Databases; https://grantsdev.cr.nps.gov/NagpraPublic/Home/Contact.
34. National Park Service, NPS National NAGPRA data related to NAGPRA Consultation Tool; https://services1.arcgis.com/fBc8EJBxQRMcHlei/arcgis/rest/services/NAGPRA_WebMap_Draft/FeatureServer.
35. USDA Forest Service, Tribal Land Cessions in the United States (2018); https://data.fs.usda.gov/geodata/edw/edw_resources/meta/S_USA.TRIBALCEDEDLANDS.xml.
36. R. Dunbar-Ortiz, *An Indigenous Peoples' History of the United States* (Beacon Press, 2014).
37. A. Greer, *Property and Dispossession: Natives, Empires, and Land in Early Modern North America* (Cambridge Univ. Press, 2018).doi: 10.1017/9781316675908
38. US Census Bureau, TIGER/Line Shapefiles and TIGER/Line Files Technical Documentation: Chapter 3-Geographic Shapefile Concepts Overview (2018); https://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2018/TGRSHP2018_TechDoc_Ch3.pdf.
39. National Weather Service, Heat Index; www.weather.gov/safety/heat-index.
40. Climatology Lab, gridMET; www.climatologylab.org/gridmet.html.
41. J. T. Abatzoglou, Development of gridded surface meteorological data for ecological applications and modeling. *Int. J. Climatol.* **33**, 121–131 (2013). doi: 10.1002/joc.3413
42. A. Dai, National Center for Atmospheric Research Staff, Palmer Drought Severity Index (PDSI), (2019); https://climatedataguide.ucar.edu/climate-data/palmer-drought-severity-index-pdsi.
43. M. Collins, R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W. J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A. J. Weaver, M. F. Wehner, M. R. Allen, T. Andrews, U. Beyerle, C. M. Bitz, S. Bony, B. B. B. Booth, in *Climate Change 2013–The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2013), pp. 1029–1136.
44. O. E. Sala, W. J. Parton, L. A. Joyce, W. K. Lauenroth, Primary Production of the Central Grassland Region of the United States. *Ecology* **69**, 40–45 (1988). doi: 10.2307/1943158
45. PRISM Climate Group, Northwest Alliance for Computational Science and Engineering, PRISM Climate Data; http://prism.oregonstate.edu.
46. 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
47. G. Dillon, "Wildfire Hazard Potential (WHP) for the conterminous United States (270-m GRID), version 2018 classified. 2nd Edition" (Forest Service Research Data Archive, 2018); www.arcgis.com/home/item.html?id=fc7f208f4bf34cf3ad34eff72261b140.
48. L. Anderson, B. Leonard, D. Parker, S. Regan, Natural resources on American Indian reservations: Blessing or curse? in *Unlocking the Wealth of Indian Nations* (Lexington, 2016), pp. 18–36.
49. B. Leonard, D. Parker, Fragmented ownership and natural resource use: Evidence from the Bakken. *Econ. J. London* **131**, 1215–1249 (2020). doi: 10.1093/ej/ueaa097
50. Maps: Oil and Gas Exploration, Resources, and Production. U.S. Energy Information Administration; https://www.eia.gov/maps/maps.htm.
51. A. A. Klingebiel, *Land-Capability Classification* (Soil Conservation Service, US Department of Agriculture, 1961).
52. N. Nunn, D. Puga, Ruggedness: The blessing of bad geography in Africa. *Rev. Econ. Stat.* **94**, 20–36 (2012). doi: 10.1162/REST_a_00161
53. L. R. Biewick, Areas of historical oil and gas exploration and production in the United States, US Geological Survey Digital Data Series DDS-69 (USGS, 2008).
54. U.S. Geological Survey, National Geospatial Intelligence Agency, Global Multi-resolution Terrain Elevation Data, 2010 (2010); www.usgs.gov/core-science-systems/eros/coastal-changes-and-impacts.
55. Congressional Research Service (CRS), "The Federal Land Management Agencies" (CRS, 2019), p. 3.
56. ProtectedLands.net, A resource for the Protected Areas Database of the United States (PAD- US); www.protectedlands.net.
57. K. Jacoby, *Crimes Against Nature: Squatters, Poachers, Thieves, and the Hidden History of American Conservation* (Univ. California Press, 2003).doi: 10.1525/9780520957930
58. M. Spence, *Dispossessing the Wilderness: Indian Removal and the Making of the National Parks* (Oxford Univ. Press, 1999). doi: 10.1093/acprof:oso/9780195142433.001.0001
59. D. Taylor, *The Rise of the American Conservation Movement: Power, Privilege, and Environmental Protection* (Duke Univ. Press, 2016). doi: 10.1215/9780822373971
60. R. Clow, The Indian Reorganization Act and the loss of tribal sovereignty: Constitutions on the Rosebud and Pine Ridge Reservations. *Great Plains Q.* **7**, 125–134 (1987).
61. V. Deloria Jr., Ed., *The Indian Reorganization Act: Congresses and Bills* (Univ. Oklahoma Press, 2002).
62. B. Duthu, Implicit divestiture of tribal powers: Locating legitimate sources of authority in Indian country. *Am. Indian Law Rev.* **19**, 353–402 (1994). doi: 10.2307/20068773
63. D. Ranco, D. Suagee, Tribal sovereignty and the problem of difference in environmental regulation: Observations on "Measured Separatism" in Indian country. *Antipode* **39**, 691–707 (2007). doi: 10.1111/j.1467-8330.2007.00547.x
64. G. Strommer, S. Osborne, The history, status, and future of tribal self-governance under the Indian Self-Determination and Education Assistant Act. *Am. Indian Law Rev.* **39**, 1–75 (2014).
65. A. Goldstein, By force of expectation: Colonization, public lands, and the property relation. *UCLA Law Rev. Discourse* **65**, 124–141 (2018).
66. V. Deloria, C. M. Lytle, *The Nations Within: The Past and Future of American Indian Sovereignty* (Univ. Texas Press, 1984).
67. B. R. Middleton, *Trust in the Land: New Directions in Tribal Conservation* (Univ. Arizona Press, 2011).
68. J. A. Shoemaker, Transforming property: Reclaiming modern indigenous land tenures. *Calif. Law Rev.* **107**, 1531–1607 (2019).
69. S. Krakoff, J.-D. Lavallee, in *Climate Change and Indigenous Peoples*, R. S. Abate, E. A. Kronk, Eds. (Edward Elgar Publishing, 2013), pp. 199–217.
70. "Legacy in Motion: Building Momentum for the Future" (2018-2019 Annual Report, National Congress of American Indians, Washington, DC, 2019); www.ncai.org/resources/ncai_publications/2018-2019-ncai-annual-report.
71. A. Parker, *Pathways to Indigenous Nation Sovereignty: A Chronicle of Federal Policy Developments* (Michigan State Univ. Press, 2018).
72. D. Wilkins, *Hollow Justice: A History of Indigenous Claims in the United States* (Yale Univ. Press, 2013).
73. L. Jantarasami, R. Novak, R. Delgado, C. Narducci, E. Marino, S. McNeeley, J. Raymond-Yakoubian, L. Singletary, K. Powys Whyte, in Impacts, Risks, and *Adaptation in the United States: Fourth National Climate Assessment*, vol. II, D. R. Reidmiller, C. W. Avery, D. R. Easterling, K. E. Kunkel, K. L. M. Lewis, T. K. Maycock, B. C. Stewart, Eds. (US Global Change Research Program, 2018), pp. 572–603.doi: 10.7930/NCA4.2018.CH15

ACKNOWLEDGMENTS

We thank M. Vernon, B. Williamson, the Native Land Information System, D. Bartecchi, A. Chesnais, D. Kane, T. Dilts, and V. Temprano. **Funding:** J.F. received support from the National Science Foundation, CAREER grant 1751483. P.B.B. was supported by the Agriculture and Food Research Initiative grant 1023124/project accession no. 1023124 from the USDA National Institute of Food and Agriculture. J.F., K.M., and P.B.B. received support from Yale University School of the Environment. K.W. received support from FORMAS #2017-01923. **Author contributions:** Conceptualization: J.F., P.B.B., K.M., and K.W. Data curation: J.F., P.B.B., K.M., J.B., G.K., and K.W. Investigation: J.B., G.K., J.F., K.M., P.B.B., and K.W. Methodology: J.B., G.K., K.M., J.F., P.B.B., and K.W. Software: J.B., G.K., and K.M. Validation: J.B., K.M., J.F., P.B.B., and K.W. Writing, original draft: J.F. Writing, review and editing: J.F., P.B.B., K.M., K.W., and J.B. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** There are no data availability restrictions. All data and reproducible code are available through the Open Science Framework at https://osf.io/3cfum and http://NativeLandResearch.org.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abe4943
Materials and Methods
Fig. S1
Tables S1 to S19
References (*74*–*82*)

27 August 2020; accepted 20 August 2021
10.1126/science.abe4943

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

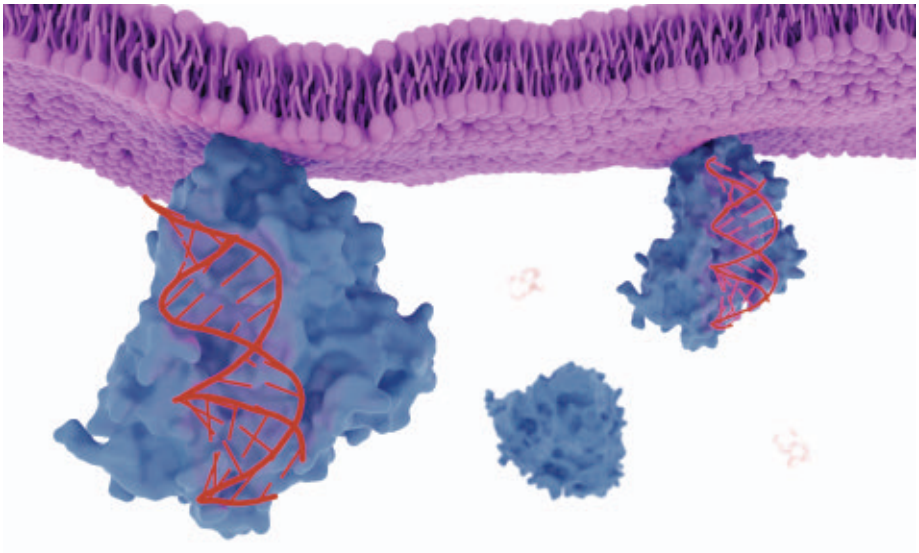
A prenylated dsRNA sensor protects against severe COVID-19

Arthur Wickenhagen, Elena Sugrue†, Spyros Lytras†, Srikeerthana Kuchi†, Marko Noerenberg†, Matthew L. Turnbull†, Colin Loney, Vanessa Herder, Jay Allan, Innes Jarmson, Natalia Cameron-Ruiz, Margus Varjak, Rute M. Pinto, Jeffrey Y. Lee, Louisa Iselin, Natasha Palmalux, Douglas G. Stewart, Simon Swingler, Edward J. D. Greenwood, Thomas W. M. Crozier, Quan Gu, Emma L. Davies, Sara Clohisey, Bo Wang, Fabio Trindade Maranhão Costa, Monique Freire Santana, Luiz Carlos de Lima Ferreira, Lee Murphy, Angie Fawkes, Alison Meynert, Graeme Grimes, ISARIC4C Investigators, Joao Luiz Da Silva Filho, Matthias Marti, Joseph Hughes, Richard J. Stanton, Eddie C. Y. Wang, Antonia Ho, Ilan Davis, Ruth F. Jarrett, Alfredo Castello, David L. Robertson, Malcolm G. Semple, Peter J. M. Openshaw, Massimo Palmarini, Paul J. Lehner, J. Kenneth Baillie, Suzannah J. Rihn, Sam J. Wilson*

INTRODUCTION: Interferons (IFNs) are cytokines that are rapidly deployed in response to invading pathogens. By initiating a signaling cascade that stimulates the expression of hundreds of genes, IFNs create an antiviral state in host cells. Because IFNs heavily influence COVID-19 outcomes, and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) replication can be inhibited by the antiviral state, it is important to understand how the individual antiviral effectors encoded by IFN-stimulated genes (ISGs) inhibit SARS-CoV-2.

RATIONALE: We hypothesized that IFN-stimulated antiviral effectors can inhibit SARS-CoV-2, and that variation at the loci encoding these defenses underlies why some people are more susceptible to severe COVID-19.

RESULTS: We used arrayed ISG expression screening to reveal that 2′-5′-oligoadenylate synthetase 1 (OAS1) consistently inhibited SARS-CoV-2 in different contexts. Using CRISPR-Cas9, we found that endogenous OAS1 makes a substantial contribution to the antiviral state by recognizing short stretches of double-stranded RNA (dsRNA) and activating RNase L. We globally mapped where OAS1 binds to SARS-CoV-2 viral RNAs and found that OAS1 binding is remarkably specific, with two conserved stem loops in the SARS-CoV-2 5′-untranslated region (UTR) constituting the principal viral target.



Location, location, location, and dsRNA sensing. A common genetic polymorphism determines whether people make a membrane-associated form of OAS1, a dsRNA sensor. Prenylation targets OAS1 to sites rich in viral dsRNA. Once in the correct place, OAS1 senses dsRNA structures in the SARS-CoV-2 5′-UTR and initiates a potent block to SARS-CoV-2 replication. OAS1 based on PDB ID 4IG8.

OAS1 expression was readily detectable at the sites of infection in individuals who died of COVID-19, and specific OAS1 alleles are known to be associated with altered susceptibility to infection and severe disease. It had previously been reported that alleles containing a common splice-acceptor single nucleotide polymorphism in OAS1 (Rs10774671) were associated with less severe COVID-19. We determined that people with at least one allele with a G at this position could express a prenylated form of OAS1 (p46), whereas other individuals could not. Using a series of mutants, we found that C-terminal prenylation was necessary for OAS1 to initiate a block to SARS-CoV-2. Furthermore, confocal microscopy revealed that prenylation targeted OAS1 to perinuclear structures rich in viral dsRNA, whereas non-prenylated OAS1 was diffusely localized and unable to initiate a detectable block to SARS-CoV-2 replication.

The realization that prenylation is essential for OAS1-mediated sensing of SARS-CoV-2 allowed us to examine the transcriptome of infected patients and investigate whether there was a link between the expression of prenylated OAS1 and SARS-CoV-2 disease progression. Analysis of the OAS1 transcripts from 499 hospitalized COVID-19 patients revealed that expressing prenylated OAS1 was associated with protection from severe COVID-19.

Because prenylated OAS1 was so important in human cases, we wanted to determine whether horseshoe bats, the likely source of SARS-CoV-2, possessed the same defense. When we examined the genomic region where the prenylation signal should reside, retrotransposition of a long terminal repeat sequence had ablated this signal, preventing the expression of prenylated anti-CoV OAS1 in these bats.

CONCLUSION: C-terminal prenylation targets OAS1 to intracellular sites rich in viral dsRNA, which are likely the SARS-CoV-2 replicative organelles. Once in the right place, OAS1 binds to dsRNA structures in the SARS-CoV-2 5′-UTR and initiates a potent block to SARS-CoV-2 replication. Thus, the correct targeting of OAS1 and the subsequent inhibition of SARS-CoV-2 likely underpins the genetic association of alleles containing a G at Rs10774671 with reduced susceptibility to infection and severe disease in COVID-19. Moreover, the conspicuous absence of this antiviral defense in horseshoe bats potentially explains why SARS-CoV-2 is so sensitive to this defense in humans. ■

The list of author affiliations is available in the full article online.
***Corresponding author. Email: sam.wilson@glasgow.ac.uk**
†These authors contributed equally to this work.
Cite this article as A. Wickenhagen *et al.*, *Science* 374, eabj3624 (2021). DOI: 10.1126/science.abj3624

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj3624

RESEARCH ARTICLE

CORONAVIRUS

A prenylated dsRNA sensor protects against severe COVID-19

Arthur Wickenhagen[1], Elena Sugrue[1]†, Spyros Lytras[1]†, Srikeerthana Kuchi[1]†, Marko Noerenberg[1]†, Matthew L. Turnbull[1]†, Colin Loney[1], Vanessa Herder[1], Jay Allan[1], Innes Jarmson[1], Natalia Cameron-Ruiz[1], Margus Varjak[1], Rute M. Pinto[1], Jeffrey Y. Lee[2], Louisa Iselin[1,2,3], Natasha Palmalux[1], Douglas G. Stewart[1], Simon Swingler[1], Edward J. D. Greenwood[4], Thomas W. M. Crozier[4], Quan Gu[1], Emma L. Davies[1], Sara Clohisey[5], Bo Wang[5], Fabio Trindade Maranhão Costa[6], Monique Freire Santana[7], Luiz Carlos de Lima Ferreira[8], Lee Murphy[9], Angie Fawkes[9], Alison Meynert[10], Graeme Grimes[10], ISARIC4C Investigators, Joao Luiz Da Silva Filho[11], Matthias Marti[11], Joseph Hughes[1], Richard J. Stanton[12], Eddie C. Y. Wang[12], Antonia Ho[1], Ilan Davis[2], Ruth F. Jarrett[1], Alfredo Castello[1], David L. Robertson[1], Malcolm G. Semple[13,14], Peter J. M. Openshaw[15,16], Massimo Palmarini[1], Paul J. Lehner[4], J. Kenneth Baillie[5,10,17], Suzannah J. Rihn[1], Sam J. Wilson[1]*

Inherited genetic factors can influence the severity of COVID-19, but the molecular explanation underpinning a genetic association is often unclear. Intracellular antiviral defenses can inhibit the replication of viruses and reduce disease severity. To better understand the antiviral defenses relevant to COVID-19, we used interferon-stimulated gene (ISG) expression screening to reveal that 2′-5′-oligoadenylate synthetase 1 (OAS1), through ribonuclease L, potently inhibits severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). We show that a common splice-acceptor single-nucleotide polymorphism (Rs10774671) governs whether patients express prenylated OAS1 isoforms that are membrane-associated and sense-specific regions of SARS-CoV-2 RNAs or if they only express cytosolic, nonprenylated OAS1 that does not efficiently detect SARS-CoV-2. In hospitalized patients, expression of prenylated OAS1 was associated with protection from severe COVID-19, suggesting that this antiviral defense is a major component of a protective antiviral response.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus responsible for the COVID-19 pandemic, first emerged in humans in 2019 and has left an indelible mark on global health, culture, and prosperity. SARS-CoV-2 is particularly sensitive to inhibition by type I interferons (IFNs), and because type I IFNs heavily influence COVID-19 outcome, there is great interest in understanding how individual IFN-stimulated genes (ISGs) inhibit SARS-CoV-2. Specifically, allelic variants of genes within the IFN system are associated with severity of COVID-19 (*1*). Moreover, neutralizing anti-IFN autoantibodies likely prevent host IFN responses from controlling SARS-CoV-2 replication (*2*), promoting severe COVID-19. Accordingly, recombinant IFNs have therapeutic potential (*3*), although the correct timing of IFN responses or the administration of recombinant IFNs is likely critical (*4*).

ISG expression screening identifies candidate anti–SARS-CoV-2 effectors

We hypothesized that variation in individual ISGs likely underlies some of the observed differential susceptibility to severe COVID-19. To identify the ISG products that inhibit SARS-CoV-2, we used arrayed ISG expression screening (*5*, *6*). We first confirmed that SARS-CoV-2 was IFN sensitive in transformed human lung A549 cells that were modified to express the SARS-CoV-2 receptor angiotensin-converting enzyme 2 (ACE2) and the serine protease transmembrane serine protease 2 (TMPRSS2) (*7*, *8*). Although these cells supported efficient viral replication, SARS-CoV-2 was potently inhibited by type I IFN treatment in this context (fig. S1, A and B). Thus, A549 cells were deemed suitable for ISG expression screening.

Because exogenously expressed ISGs can trigger antiviral gene expression pathways (*5*), we used transformed ACE2-expressing and interferon regulatory factor 3 (IRF3)–deficient A549 cells (A549-Npro-ACE2), which have an attenuated ability to produce IFN (*9*), as the background for the screens. We transduced these cells with an arrayed library of lentiviral vector–encoded ISGs in a 96-well plate format (one ISG per well) using a library of >500 human ISGs and a similar library of >350 rhesus macaque ISGs (*5*) (fig. S1, C to E). The macaque ISGs were included because they increased the total number of ISGs under consideration (including orthologs and additional ISGs). About two-thirds of the ISGs examined could potentially be relevant to betacoronavirus infection (*10*) (fig. S1, F and G). To capture inhibition that might occur at different stages of the virus lifecycle, we used a green fluorescent protein (GFP)–encoding recombinant SARS-CoV-2 (*11*) and measured the ability of each individual ISG to inhibit SARS-CoV-2 at 14 hours (early) and 40 hours (late) after infection. Using this approach, we identified several candidate anti–SARS-CoV-2 effectors (Fig. 1A). All ISGs that conferred more than twofold inhibition at early and late time points, or only at late time points, were considered potential "candidates" and underwent further independent confirmatory "miniscreens." The magnitude of protection conferred by each candidate ISG at early and late time points was assessed using ACE2-positive cells in the presence or absence of IRF3 (Fig. 1B and fig. S1, H to L). In addition, we sought to subtract nonspecific inhibitors of SARS-CoV-2 by identifying ISGs that triggered a polygenic antiviral state (Fig. 1C) or caused cytotoxicity (Fig. 1D). After these confirmatory and negative selection screens, we identified six candidate antiviral effectors that robustly inhibited SARS-CoV-2 without inducing substantial toxicity or inducing interferon-stimulated response element (ISRE) expression.

These candidate effectors included known antiviral genes such as the IFN-inducible short isoform (isoform 4) of NCOA7, which inhibits influenza A viruses (IAVs) (*12*), and 2′-5′-oligoadenylate synthetase 1 (OAS1), a double-stranded RNA (dsRNA) sensor capable of activating ribonuclease L (RNase L) (*13*, *14*). We also identified UNC93B1, a polytopic membrane protein required for TLR trafficking (*15*), as well as SCARB2, a virus receptor (*16*) involved in cholesterol transport (*17*, *18*). In addition, we identified ANKFY1 and ZBTB42,

[1]Medical Research Council–University of Glasgow Centre for Virus Research (CVR), Institute of Infection, Inflammation and Immunity, University of Glasgow, Glasgow, UK. [2]Department of Biochemistry, University of Oxford, Oxford, UK. [3]Nuffield Department of Medicine, University of Oxford, Oxford, UK. [4]Cambridge Institute of Therapeutic Immunology and Infectious Disease, University of Cambridge, Cambridge, UK. [5]Roslin Institute, University of Edinburgh, Edinburgh, UK. [6]Laboratory of Tropical Diseases, Department of Genetics, Evolution, Microbiology and Immunology, Institute of Biology, University of Campinas, Campinas, Sao Paolo, Brazil. [7]Department of Education and Research, Oncology Control Centre of Amazonas State (FCECON), Manaus, Amazonas, Brazil. [8]Postgraduate Program in Tropical Medicine, Tropical Medicine Foundation Dr. Heitor Vieira Dourado, Manaus, Amazonas, Brazil. [9]Edinburgh Clinical Research Facility, University of Edinburgh, Western General Hospital, Edinburgh, UK. [10]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK. [11]Wellcome Centre for Molecular Parasitology, Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, UK. [12]Division of Infection & Immunity, Cardiff University, Cardiff, UK. [13]NIHR Health Protection Research Unit for Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK. [14]Respiratory Medicine, Alder Hey Children's Hospital, Liverpool, UK. [15]National Heart and Lung Institute, Imperial College London, London, UK. [16]Imperial College Healthcare, National Health Service Trust London, London, UK. [17]Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, UK.
*Corresponding author. Email: sam.wilson@glasgow.ac.uk
†These authors contributed equally to this work.

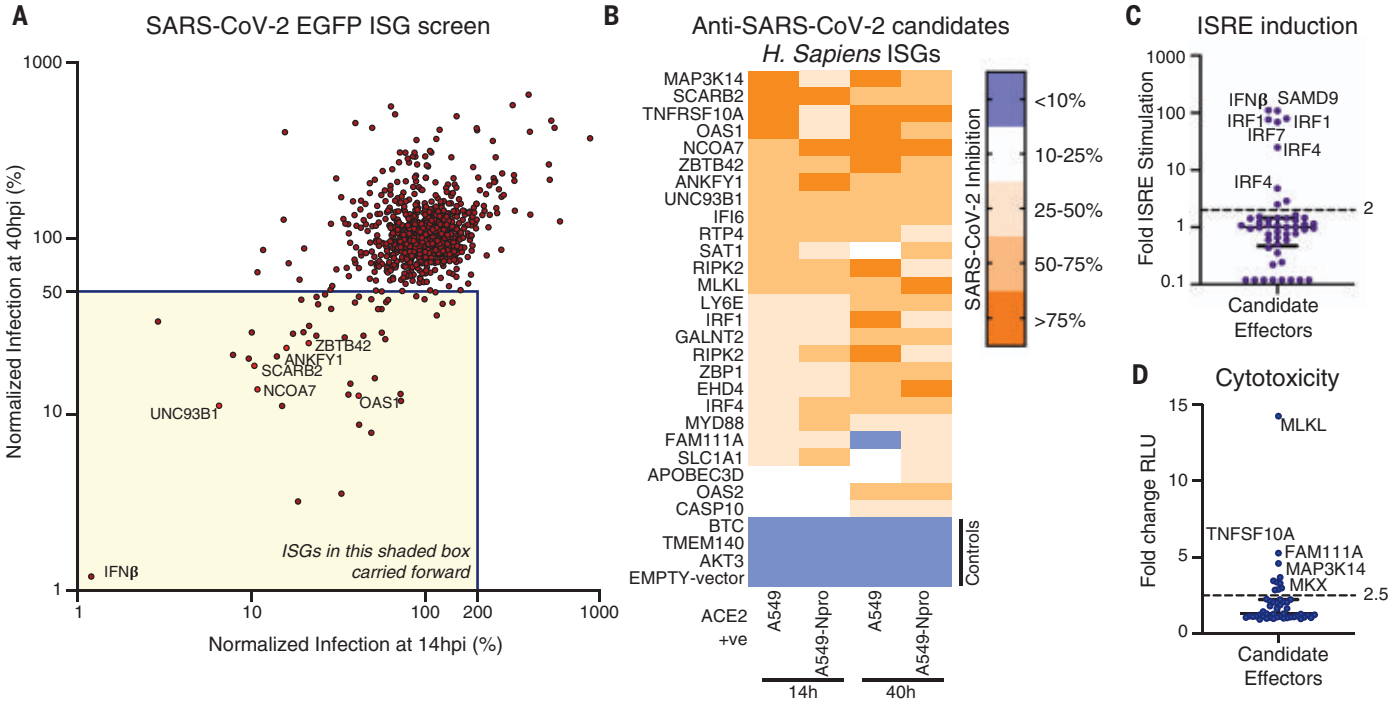


Fig. 1. Arrayed ISG expression screening reveals factors with candidate anti–SARS-CoV-2 activity. (**A**) A549-Npro-ACE2 cells were transduced with hundreds of individual human or macaque ISGs (see fig. S1, C to E) and infected with SARS-CoV-2-EGFP (Wuhan-1) in duplicate, and the level of infection in the presence of each ISG was measured using flow cytometry at 14 and 40 hours postinfection (hpi). (**B**) Miniscreen of the ability of human candidate effectors identified in (A) alongside controls to inhibit SARS-CoV-2 in A549 and A549-Npro at 14 and 40 hpi (the equivalent panel for macaque ISGs presented in fig. S1L). (**C** and **D**) The ability of each human and macaque effector to either stimulate ISRE activity using A549-ISRE-EGFP cells (C) or cause toxicity (CytoTox-Glo) using supernatant from the same A549-ISRE-EGFP cells (D) at 48 hours after transduction with the relevant ISG-encoding lentiviral vector. The dashed line indicates threshold for negative selection.

which have not previously been ascribed antiviral activity. We exogenously expressed these ISGs in human A549 lung cells modified to express either ACE2 or both TMPRSS2 and ACE2, and examined their ability to inhibit an isolate of SARS-CoV-2 [CVR-GLA-1 isolated in March 2020 (*8*)]. In the absence of TMPRSS2, all six candidate antiviral effectors inhibited SARS-CoV-2 (2.5- to >1000-fold) (Fig. 2A), whereas only OAS1 consistently inhibited this "early" isolate in different cell backgrounds, irrespective of TMPRSS2 status (Fig. 2, A to C). Indeed, TMPRSS2-mediated entry has been proposed as a strategy used by SARS-CoV-2 to evade antiviral factors (*19*), including NCOA7 (*20*). Moreover, the initial screens were executed in the absence of TMPRSS2 (Fig. 1) potentially biasing our "hit list" toward effectors that might be inactive in TMPRSS2-expressing cells.

OAS1 exhibits multiple characteristics expected from an anti–SARS-CoV-2 effector

To identify effectors present at the sites of SARS-CoV-2 infection, we examined the IFN responsivity of the six candidate effectors using published studies from the interferome version 2.01 database (Fig. 2D) (*21*). We also determined their basal expression in respiratory and gastrointestinal tracts (GTEx) (Fig. 2E) and assessed their transcript abundance in postmortem lung tissue from COVID-19 patients (Fig. 2F and fig. S2A). In addition, we examined the genomic locus of each candidate effector for single nucleotide polymorphisms (SNPs) within alleles that may be associated with increased susceptibility to infection and/or severe disease (Fig. 2G and fig. S2B). Consistent with a potential role in influencing susceptibility to SARS-CoV-2 infection (*22*), OAS1 was frequently detected in nasal epithelium [which is a potential site of initial infection (*23*)] sampled from healthy individuals (three of six individuals) (fig. S2C). After these analyses, we focused our attention on OAS1 because its antiviral activity was the most robust (Fig. 2, A to C), its low basal transcription was highly IFN inducible (Fig. 2, D and E), its mRNA was readily detectable in infected patients (Fig. 2F), and common allelic variants were associated with altered susceptibility to infection and severe disease (Fig. 2G) (*1*, *22*, *24*, *25*).

The block initiated by OAS1 is not dependent on OAS3

OAS1 is an evolutionarily ancient ISG that has maintained IFN responsivity for hundreds of millions of years (*26*). The OAS system was one of the first antiviral pathways to be defined (*27*), and the canonical model of OAS antiviral activity involves initial dsRNA sensing by OASs, which results in the synthesis of 2′-5′-linked oligoadenylates (2-5A). 2-5A induces the dimerization of inactive RNase L, which then mediates the indiscriminate cleavage of viral and host RNAs presenting single-stranded UpU and UpA motifs (*28*). The initial sensing of virus dsRNA that subsequently activates RNase L has mostly been ascribed to OAS3, with OAS1 infrequently considered as a major viral dsRNA sensor (*29*). Indeed, in >30 arrayed ISG screens completed in our laboratory, SARS-CoV-2 was the only virus substantially inhibited by OAS1 (Fig. 3A). We therefore investigated whether OAS1 antiviral activity was dependent upon OAS3. OAS3 is readily detectable in A549 cells (Fig. 3B), and additional exogenous OAS3 had no effect on the replication of SARS-CoV-2 (Fig. 3, B and C). Removal of OAS3 to undetectable levels using CRISPR-Cas9 did not attenuate the ability of exogenous OAS1 to instigate a block to SARS-CoV-2 (Fig. 3D). Similarly, exogenous expression of OAS1 initiated a block to SARS-CoV-2 in HT1080 cells, which have low or undetectable levels of basal OAS expression

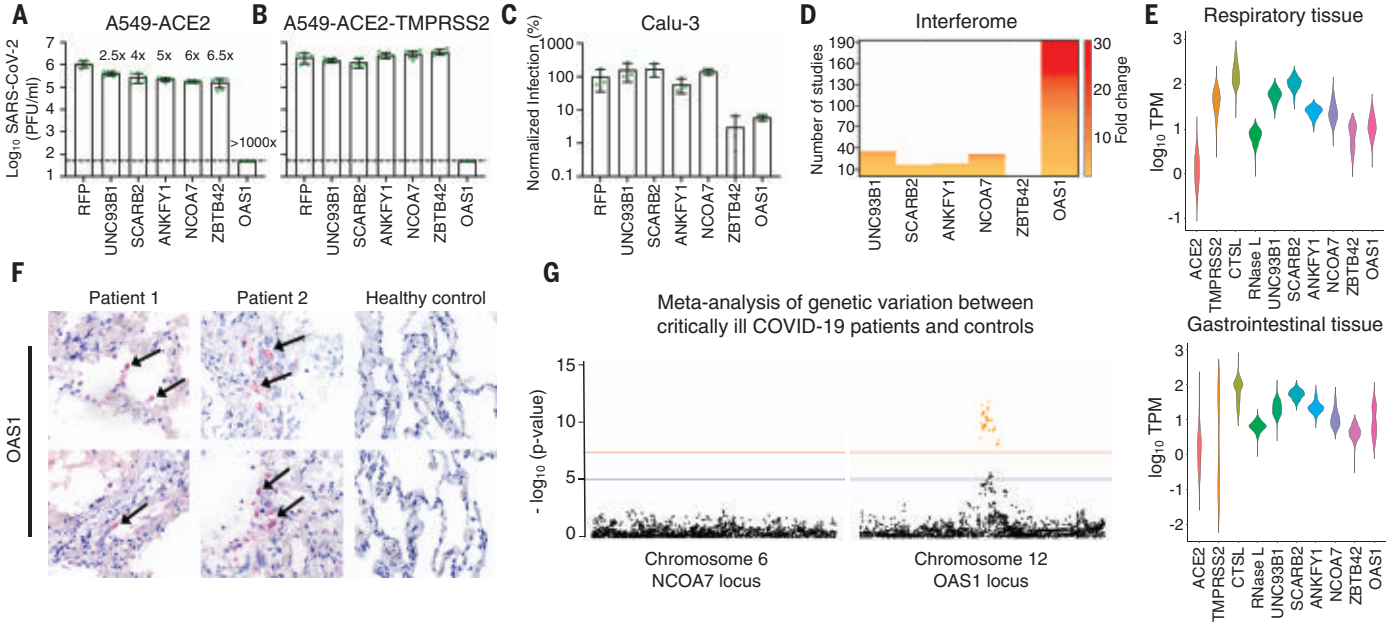


Fig. 2. The ISG OAS1 initiates a block to SARS-CoV-2 replication. (**A** and **B**) SARS-CoV-2 isolate CVR-GLA-1 infectious titers (PFU/ml) were determined using A549-ACE2 (A) or A549-ACE2-TMPRSS2 cells (B) modified to express the candidate effectors (UNC93B1, SCARB2, ANKFY1, NCOA7, ZBTB42, or OAS1) from the screening pipeline (Fig. 1, A to D). Fold protection from SARS-CoV-2 is indicated for each gene in (A). (**C**) SARS-CoV-2-ZsGreen infectious titers on Calu-3 cells expressing the same hit ISGs as in (A) and (B), measured by flow cytometry at 40 hpi. (**D**) The "ISG-ness" of selected genes was assessed by fold change upon type I IFN stimulation as reported in studies in the Interferome v2.01 database (http://www.interferome.org/). (**E**) Gene expression analysis across different respiratory and gastrointestinal tissues using datasets from the GTEx database, with ACE2 and TMPRSS2 included for reference. RNase L is included as functionally linked to OAS1. (**F**) Detection of OAS1 gene expression by RNAscope in FFPE lung tissue of deceased COVID-19 patients compared with healthy control lung tissue. Arrows indicate staining of positive cells. (**G**) Meta-analysis of the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/) for genetic variation between critical ill COVID-19 patients and control populations at the gene locus of the NCOA7 and OAS1 genes. The red line indicates the threshold for significant SNPs (yellow dots).

(Fig. 3, E and F). Thus, OAS3 is not required for OAS1 to instigate a block to SARS-CoV-2 replication. To confirm that OAS1 was inhibiting SARS-CoV-2 through the synthesis of 2-5A and activated RNase L, we next disrupted the RNase L locus in OAS1-expressing cells. The antiviral activity of OAS1 was only effective in the presence of RNase L, and the loss of RNase L abrogated the ability of OAS1 to inhibit SARS-CoV-2 (Fig. 3G). RNase L activation can, in principle, inhibit viruses by degrading viral or host RNAs (*30*, *31*), eventually resulting in apoptosis (*32*), or by triggering an IFN response (*33*). We therefore examined the contribution of RNase L–induced IFN responses to the inhibition of SARS-CoV-2 by OAS1 by ablating Janus kinase (JAK)–signal transducer and activator of transcription (STAT) signaling using the JAK inhibitor ruxolitinib (Rux) (*34*). Type I IFN treatment potently inhibited SARS-CoV-2 replication (Fig. 3H), and this effect was entirely reversed by the addition of Rux. OAS1 potently inhibited SARS-CoV-2 in the absence of JAK-STAT signaling (Fig. 3H), indicating that the RNase L–mediated destruction of host and/or viral RNAs is likely the predominant mechanism through which OAS1 inhibits SARS-CoV-2.

OAS1 senses conserved dsRNA structures in the SARS-CoV-2 5′-untranslated region

To understand how OAS1 senses SARS-CoV-2 infection, we applied individual nucleotide-resolution cross-linking and immunoprecipitation (iCLIP2) (*35*) to SARS-CoV-2–infected AAT cells. To maximize the viral RNA available to OAS1, we did this in AAT cells modified to express exogenous OAS1, which were also devoid of substantial RNase L activity (guide 5, Fig. 3G). The iCLIP approach freezes protein-RNA interactions using ultraviolet cross-linking, followed by RNase trimming to generate protein-protected RNA fragments. OAS1 can subsequently be immunoprecipitated and the cross-linked RNA reverse transcribed and sequenced. Because the amino acids cross-linked to the RNA cause termination of reverse transcription, iCLIP2 provides a single-nucleotide-resolution map of protein-binding sites within RNA molecules. We used control immunoglobulin G (IgG) immunoprecipitation and size-matched input controls (*36*) to subtract confounding sequences not derived from OAS1 binding. iCLIP2 revealed that OAS1 interacted with several regions of the SARS-CoV-2 genome (table S1), with the most prominent sites mapping to the first 54 nucleotides of the 5′-untranslated region (UTR) (*37*) that is present in all SARS-CoV-2 positive-sense viral RNAs (Fig. 3I). No substantial traces of binding were observed in the negative strand, suggesting that OAS1 likely bound positive-sense viral transcripts (as opposed to replication intermediates). The major viral target encompassed stem loops 1 and 2 (SL1 and SL2) within the 5′-UTR, consistent with the known capacity of OAS1 to interact with short regions of dsRNA (*13*). Unexpectedly, we observed a substantial enrichment of an off-template G upstream of the first nucleotide of the 5′-UTR, which is compatible with the 7-methylguanosine cap structure previously observed with the antiviral cap-binding protein GEMIN5 (fig. S3) (*38*).

To further understand OAS1-mediated sensing, we assessed the transcriptome-wide binding of OAS1. The number of host transcripts associated with OAS1 was substantially higher in infected cells compared with mock controls (Fig. 3, J and K). This suggests that OAS1 RNA-binding activity may be enhanced by SARS-CoV-2 infection. OAS1 interacted primarily with cellular RNAs that are highly structured, including small nucleolar RNAs, long noncoding RNAs, and the intronic regions of mRNAs (Fig. 3J). These data are compatible with the notion that OAS1 senses short stretches of dsRNA that are found in stem loops. We also

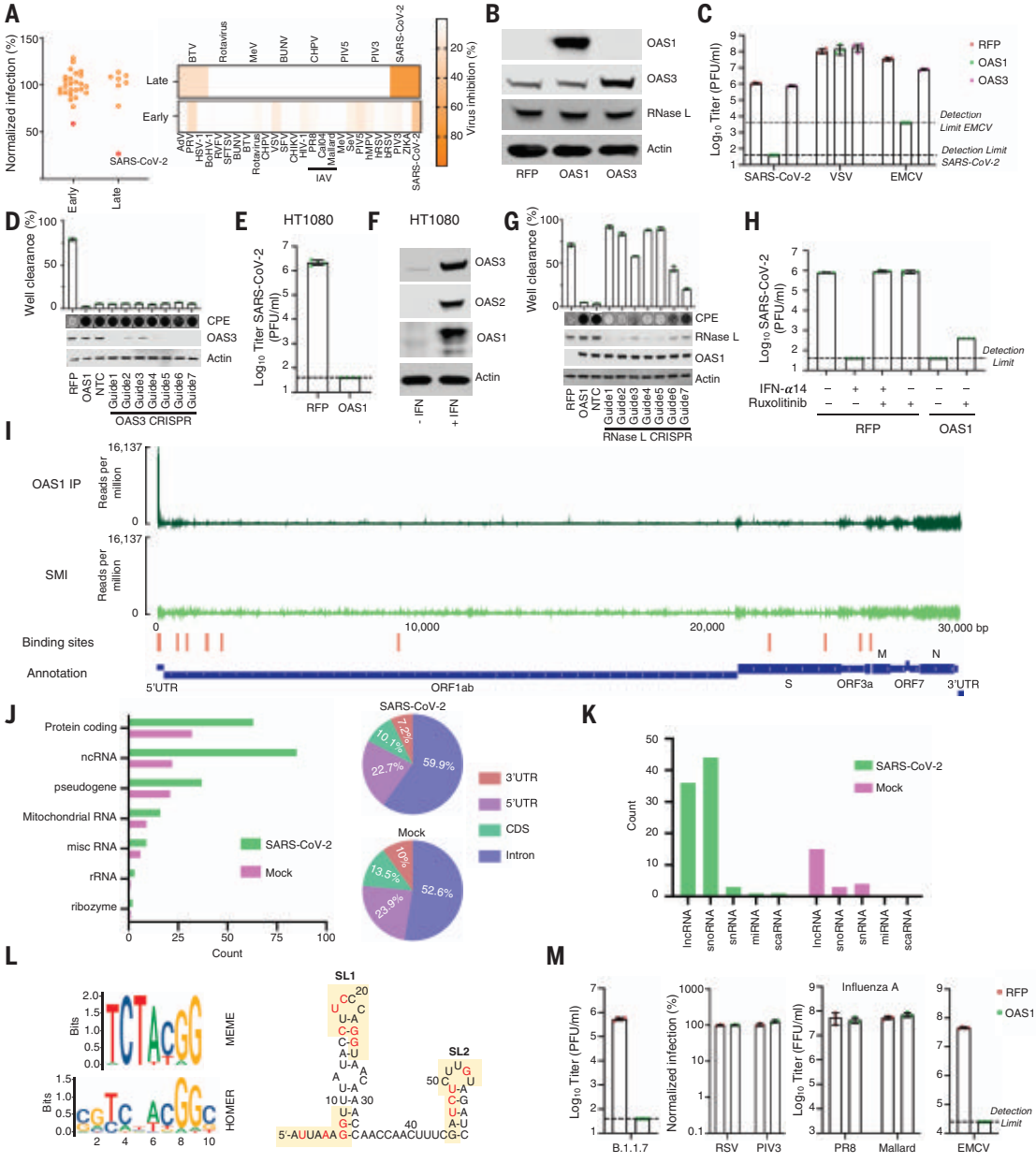


Fig. 3. OAS1 inhibition of SARS-CoV-2 is specific and mediated through the RNase L pathway. (**A**) Normalized infection in the presence of OAS1 at early or late stages of the viral life cycle, quantified in large-scale ISG expression screens (similar to Fig. 1) for a panel of viruses (described in arrayed ISG expression screening methods). (**B**) AAT cells were modified to express OAS1 and OAS3 and protein expression (OAS1, OAS3, and RNase L) in the cell lines, which was monitored using Western blotting. (**C**) The titers of SARS-CoV-2, VSV and EMCV were determined (PFU/ml) in the presence of each ISG in the cell lines characterized in (B). (**D**) SARS-CoV-2 replication (well clearance at 72 hpi due to cytopathic effects of virus replication) was assessed in AAT cells that were modified to express exogenous OAS1, and the OAS3 expression of which was reduced using seven different lentiviral vector-derived CRISPR guides and one NTC. The level of OAS3 KO was assessed by Western blotting, and a typical virus-induced CPE is shown. (**E**) SARS-CoV-2 infectious titer (PFU/ml) on HT1080-ACE2-TMPRSS2 (HAT) cells expressing RFP or OAS1. (**F**) Protein expression (OAS3, OAS2, and OAS1) in HAT cells in the presence and absence of pretreatment with 1000 pg/ml IFNα14, monitored by Western blotting. (**G**) SARS-CoV-2 replication (well clearance at 72 hpi due to cytopathic effects of virus replication) was assessed in cells with RNase L expression that was reduced using seven different lentiviral vector-derived CRIPSR guides and one NTC. The level of RNase L KO was assessed by Western blotting, and a typical virus-induced CPE is shown. (**H**) SARS-CoV-2 infectious titer (PFU/ml) on AAT cells expressing RFP or OAS1 was determined in the presence and absence of pretreatment with 100 U/ml IFNα14 and/or 0.5 μM ruxolitinib. (**I**) iCLIP2 analysis of OAS1 binding sites on SARS-CoV-2 RNA. Coverage of three replicate tracks overlaid mapped to the SARS-CoV-2 genome in the OAS1-IP and an SMI allow detection of OAS1-binding sites shown in red above the SARS-CoV-2 genome annotation. (**J**) Transcriptome-wide profiling of OAS1 iCLIP2 targets by gene biotypes in SARS-CoV-2 or mock infected cells. Pie charts indicate distribution of OAS1-binding sites within each transcript feature of protein-coding genes. (**K**) Detailed representation of OAS1 iCLIP2 targets in noncoding RNA biotypes between SARS-CoV-2 and mock infected samples. (**L**) Motif prediction of OAS1-binding sites in cellular transcripts using MEME or HOMER. Presence of these predicted binding motifs in the SL1 and SL2 loops of the 5′-UTR of SARS-CoV-2 (*37*). (**M**) AAT cells were modified to express RFP or OAS1. The infectious titer of SARS-CoV-2 B.1.1.7 and EMCV on these cells was determined by plaque assay. Similarly, titers of RSV-GFP and PIV3-GFP were determined using flow cytometry (24 hpi). Titers of influenza A viruses (IAV/H1N1/PR8 and IAV/H1N1/Mallard) were determined using an immunostained focus-forming assay.

observed that OAS1 interacted with several mitochondrial RNAs (fig. S3), which is consistent with the presence of dsRNA in the mitochondria that can induce innate immune responses (*39*).

To identify potential drivers of the high specificity of OAS1, we analyzed all binding sites present in cellular RNAs using a variety of approaches. Both MEME (*40*) and HOMER (*41*) consistently identified a prominent motif (UCUACGG), and HOMER detected two additional nucleotide signatures before (CG) and after (C) this motif. Although the first U and GG were relatively conserved, the middle of this motif was more variable (Fig. 3L). Three versions of the motifs identified in the host transcriptome were apparent in the 5′-UTR of SARS-CoV-2 RNA, suggesting that OAS1 binds SL1 and SL2 simultaneously or that multiple OAS1 molecules bind to multiple sites in this region, either on the same or distinct RNAs (Fig. 3L). Because OAS1 is known to interact with dsRNA (*13*), we also applied GLAM2 (*40*), which allows identification of binding sites with gaps. The motifs identified by GLAM2 were longer and more heterogeneous. However, it was possible to distinguish two additional features, a UU…UG and a GA…AT (fig. S3E). These gapped motifs are similar to those previously identified in RNA substrates (*42*, *43*). Indeed, the WWN_9WG motif (*43*) is also present in SL1 and SL2, with the G equivalent to position 9 in Fig. 3L. When combined, structural context and sequence specificity could potentially explain the high degree of specificity exhibited by OAS1, which is likely necessary to avoid the inappropriate activation of RNase L.

The antiviral activity instigated by OAS1 is highly specific

To further understand the specificity of OAS1-mediated sensing, we considered the ability of OAS1 to initiate inhibition of a panel of viruses that replicate using dsRNA intermediates within different subcellular compartments. We first confirmed that OAS1 was active against the more transmissible B.1.1.7 variant of SARS-CoV-2, which remained highly sensitive to OAS1 restriction (Fig. 3M). When we examined three negative-sense single-stranded RNA (ssRNA) viruses in which genome replication occurs in the cytosol [Indiana vesiculovirus (VSV), human respirovirus 3 (PIV-3), and human respiratory syncytial virus (RSV)], all were unaffected by OAS1 (Fig. 3, C and M). Similarly, influenza A viruses (which have a segmented negative-sense ssRNA genome), the replication of which occurs in the nucleus, were completely resistant to OAS1 (Fig. 3M). By contrast, when we examined cardiovirus A (EMCV), a positive-sense ssRNA virus in which genome replication occurs within replicative organelles and double-membrane vesicles (*44*), we found that OAS1 restricted this virus by >100-fold (Fig. 3, C and M) (*45*). This notable antiviral specificity could be driven by virus-specific OAS evasion or antagonism strategies (*46*), such as the evasion role proposed for IAV NS1 (*47*). However, because CoVs replicate in similar endoplasmic reticulum–derived membranous structures to EMCV (*48*–*50*), we also considered whether OAS1 might be a dsRNA sensor specifically targeted to membranous replicative organelles in infected cells.

C-terminal prenylation is necessary for OAS1 to initiate a block to SARS-CoV-2

In humans, the OAS1 protein is expressed as two major forms designated p46 and p42. The longer p46 isoform (present in the screening library and used in Figs. 1 to 3) is generated by alternative splicing to an exon downstream of the terminal exon used by the p42 isoform (Fig. 4, A and B). Although all human genotypes contain the exon that completes the transcript encoding p46, an intronic SNP (Rs10774671, also known as 12-112919388-G-A) determines OAS1 exon usage. Alleles with a G at this SNP (G alleles) specify expression of the p46 isoform and some p42, whereas alleles with A at this position predominantly encode the p42 isoform and cannot express the p46 isoform (*51*–*53*). Individuals with G alleles are more resistant to West Nile virus infection (*54*) and respond better to IFN therapy after hepatitis C infection (*55*). G alleles are also associated with protection against severe COVID-19 disease (*1*, *22*, *24*, *25*). We thus investigated whether p42 has the same anti–SARS-CoV-2 activity as p46. The p42 isoform, which is the most common isoform in humans (~61% of alleles), had no detectable anti–SARS-CoV-2 activity (Fig. 4C). Although differential basal enzymatic activity was initially proposed to underlie the divergent antiviral potential of p46 and p42 OAS1 (*51*), this effect is likely due to expression level because p42 and p46 have similar catalytic activities (*56*). Examination of the C terminus of p46 reveals that it encodes a canonical CAAX-box prenylation signal (CTIL) that is absent from the p42 variant (Fig. 4B) and is predicted to be geranylgeranylated (*57*). Indeed, prenylation of OAS1 was proposed to alter the subcellular localization of OAS1, perhaps influencing mitochondrial respiration (*58*). We therefore hypothesized that prenylated OAS1 is targeted to membranous viral replicative organelles and facilitates the sensing of CoV dsRNA (and perhaps many divergent picornaviruses, arteriviruses, caliciviruses, and flaviviruses that also use replicative organelles). To test this, we introduced a point mutation into the p46 isoform to prevent its prenylation (C397A), and this completely ablated the antiviral activity of p46 (Fig. 4C). Similarly, appending a four-amino-acid CAAX-box (CTIL) to the C terminus of the p42 isoform conferred substantial antiviral activity to the inactive p42 form, reducing the ability of SARS-CoV-2 to form plaques by >100-fold (Fig. 4C). Thus, prenylation of OAS1 appears to be necessary for dsRNA sensing of SARS-CoV-2. A nearly identical picture emerged using EMCV (Fig. 4D). Although prenylated p46 and p42-CTIL reduced EMCV plaque formation by >100-fold, nonprenylated p42 or p46 C397A had no anti-EMCV activity. This antiviral activity of p46 and p42-CTIL was highly specific and did not inhibit the ability of VSV to form plaques in parallel experiments (fig. S4A). Although potent, prenylated p42-CTIL instigated a weaker block to SARS-CoV-2 than p46 (Fig. 4C), we therefore considered whether additional determinants resided in the C-terminal region. A relatively short (18-aa) fragment of the p46 C terminus (the same length as p42-CTIL) was indistinguishable from p46 in its ability to initiate a block to SARS-CoV-2 (Fig. 4E), indicating that most of the 54-aa C terminus of p46 is dispensable for efficient inhibition of SARS-CoV-2.

Endogenous OAS1 makes a substantial contribution to the antiviral state

To evaluate the contribution that OAS1 makes to the antiviral state, we ablated OAS1 expression in HT1080-ACE2-TMPRSS2 cells, which are heterozygous at Rs10774671 and predominantly express p46 after IFN stimulation (Fig. 4F) (*29*). We then ablated OAS1 expression using CRISPR-Cas9 and examined the ability of type I IFN to inhibit EMCV replication (Fig. 4F). Notably, IFN pretreatment inhibited ~20 times more virus replication in the presence of OAS1 (Fig. 4G). In the absence of OAS1 expression, about seven times as much IFN was required to inhibit viral replication than when OAS1 was present [half-maximal inhibitory concentration (IC_{50}) of 25 pg/ml in the control versus an IC_{50} of 188 pg/ml in the OAS1 knock-out (KO)] (Fig. 4H). Thus, the antiviral inhibition initiated by endogenous OAS1 can play a major role in generating an antiviral state. Similar unpublished experiments with SARS-CoV-2 were attempted but were less conclusive, likely because SARS-CoV-2 was so profoundly inhibited by type I IFNs. Indeed, IFN treatment reduced viral replication by several orders of magnitude in A549 cells that cannot express prenylated OAS1 (*29*) (fig. S1). This indicates that in the IFN-stimulated cell, prenylated OAS1 is likely only one of multiple ISGs that mediate the potent inhibition of SARS-CoV-2.

Prenylated OAS1 colocalizes with viral dsRNA

Because SARS-CoV-2 replication, which uses dsRNA intermediates, occurs within membranous replicative organelles, we next considered whether prenylation localizes OAS1 to these replicative organelles. Prenylated p46 and p42-CTIL localized to membranous perinuclear

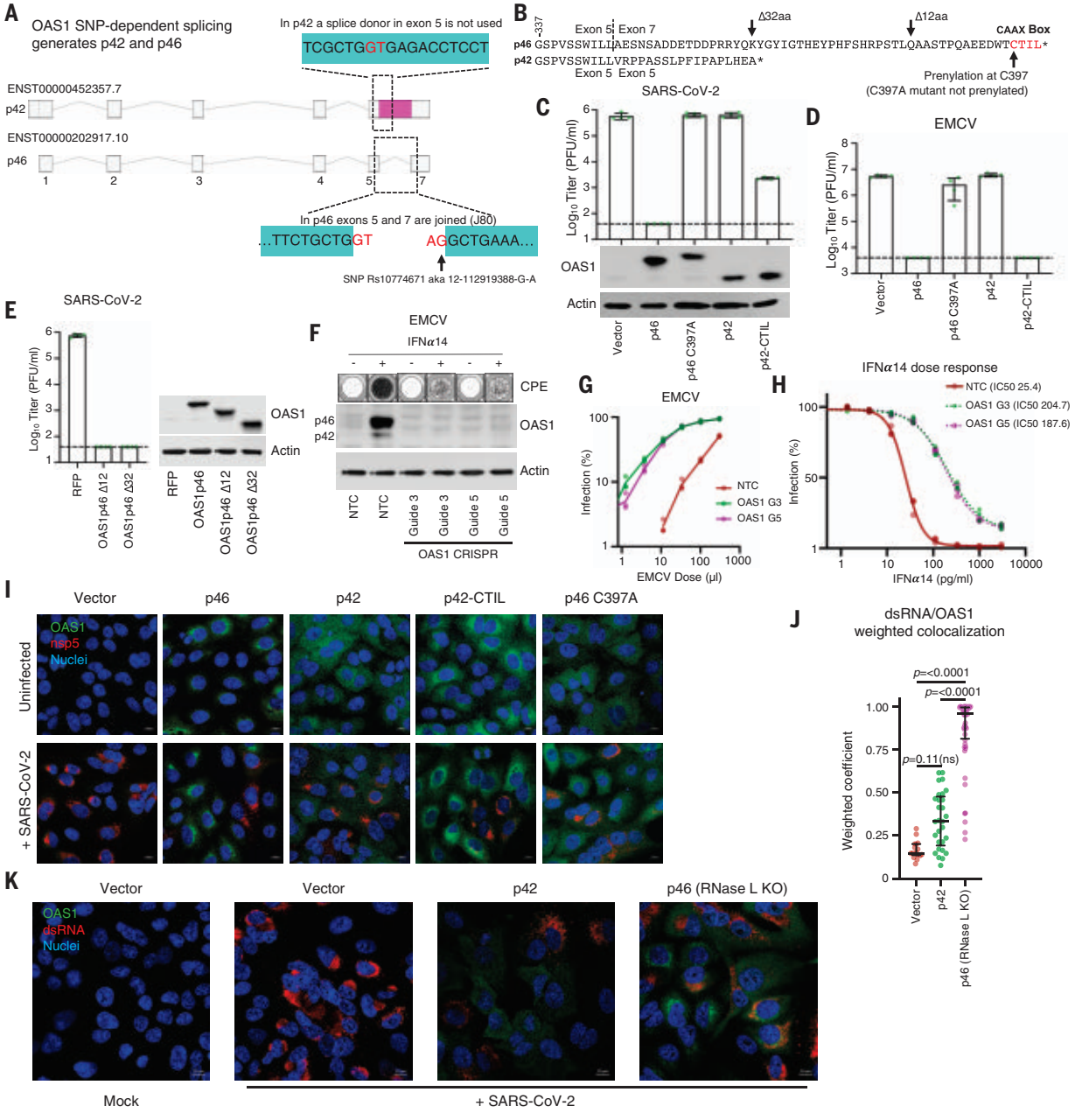
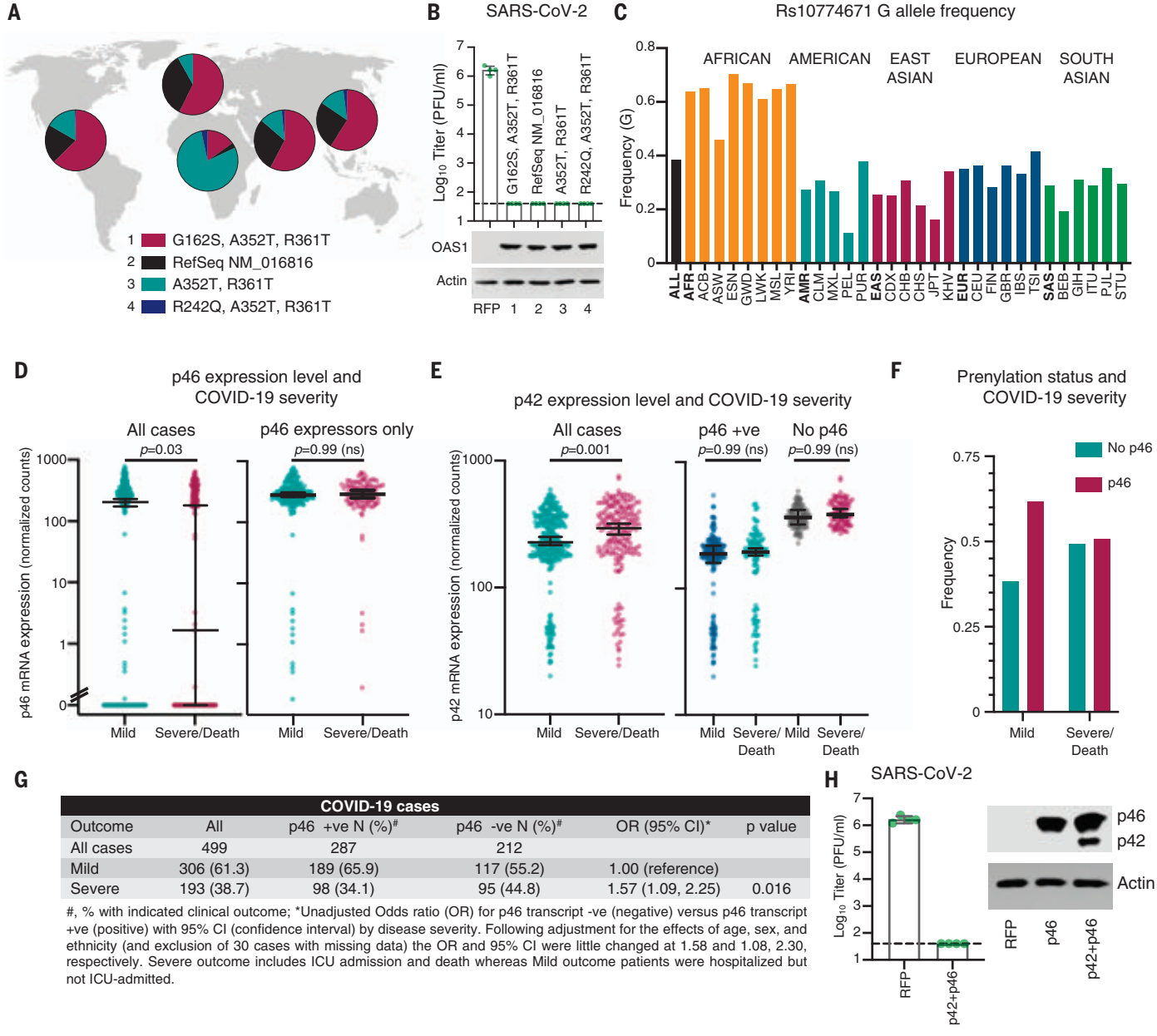


Fig. 4. OAS1 isoforms have differential antiviral activity as determined by C-terminal prenylation. (**A**) Schematic representation of OAS1 splicing resulting in isoforms p42 and p46. The area shaded in pink is exonic in p42 and intronic in p46. (**B**) Protein sequence alignment of the p46 and p42 isoforms, indicating the CAAX box prenylation signal in p46 and locations of modifications made in this work. (**C**) SARS-CoV-2 infectious titer (PFU/ml) on AAT cells expressing the OAS1 isoforms p46, not prenylated p46 (p46 C397A), p42 or prenylated p42 (p42CTIL) or a vector control. Protein expression analysis of the levels of isoforms and mutants is shown by Western blot. (**D**) EMCV infectious titer on the cells from (B) as determined by plaque assay (PFU/ml). (**E**) SARS-CoV-2 infectious titer (PFU/ml) on AAT cells expressing OAS1 p46 or the p46 C-terminal truncations OAS1 p46 Δ12 and OAS1 p46 Δ32. The level of expression is shown by Western blotting. (**F**) EMCV replication in HAT cells with reduced OAS1 expression using two different lentiviral vector-derived CRISPR guides and one NTC guide. Well clearance at 24 hpi was assessed in the presence or absence of pretreatment with 1000 pg/ml IFNα14 (typical wells are shown in the top panel), and the level of OAS1 KO was assessed by Western blotting. (**G**) EMCV infectious virus titration (based on percentage well clearance) in HAT cells in which OAS1 expression was reduced using two different OAS1 KO guides compared with a NTC. (**H**) EMCV infection (percentage well clearance) after pretreatment of various doses of IFNα14 in same cells as in (G). (**I**) Representative immunofluorescence on cells from (C) infected with SARS-CoV-2 isolate CVR-GLA-1 at MOI 0.5 for 24 hours, followed by staining with anti-OAS1 (green) and anti–SARS-CoV-2-nsp5 (red) antibodies, and nuclear Hoechst stain (blue). Contrast was reduced in the p46 sample to prevent oversaturation in the green channel caused by particularly strong perinuclear concentration. Representative cells from one out of three independently performed experiments are depicted. (**J**) Quantification of colocalization of dsRNA with OAS1 (weighted colocalization coefficient) in infected cells represented in (K). Each data point represents a distinct region of interest encompassing an individual cell from one representative experiment. (**K**) Representative immunofluorescence on AAT cells modified with a vector control, OAS1 p42, or OAS1 p46 in the presence of RNase L KO, infected or mock treated with SARS-CoV-2 isolate GLA-1 at MOI 0.5 for 24 hours, followed by staining with anti-OAS1 (green) and anti-dsRNA (red) antibodies and nuclear Hoechst stain (blue). Representative cells from one out of two independently performed experiments are depicted.



G

COVID-19 cases					
Outcome	All	p46 +ve N (%)#	p46 -ve N (%)#	OR (95% CI)*	p value
All cases	499	287	212		
Mild	306 (61.3)	189 (65.9)	117 (55.2)	1.00 (reference)	
Severe	193 (38.7)	98 (34.1)	95 (44.8)	1.57 (1.09, 2.25)	0.016

#, % with indicated clinical outcome; *Unadjusted Odds ratio (OR) for p46 transcript -ve (negative) versus p46 transcript +ve (positive) with 95% CI (confidence interval) by disease severity. Following adjustment for the effects of age, sex, and ethnicity (and exclusion of 30 cases with missing data) the OR and 95% CI were little changed at 1.58 and 1.08, 2.30, respectively. Severe outcome includes ICU admission and death whereas Mild outcome patients were hospitalized but not ICU-admitted.

Fig. 5. Prenylated OAS1 protects against severe COVID-19. (**A**) Allelic frequencies of the most common circulating p46 variants of OAS1 displayed by region. (**B**) Infectious titers of SARS-CoV-2 CVR-GLA-1 (PFU/ml) were determined on AAT cells modified to express each human p46 OAS1 variant. OAS1 expression was monitored using Western blotting (lower panels). (**C**) Frequency of alleles with G at Rs10774671 in different human populations (1000 Genomes Project). The population names are expanded in the materials and methods. (**D**) Transcript abundance of the p46 isoform (encoding prenylated OAS1), determined using JunctionSeq analysis (J080) of RNA-seq data from whole blood from infected patients with mild (hospitalized but not ICU admitted) or severe or lethal (ICU admitted and/or death) COVID-19. (**E**) Transcript abundance of the p42 isoform (E037) determined as in (E). For (D) and (E), significance was determined using a Mann–Whitney *U* test except where multiple comparisons were made [(E), right], and then a Kruskal-Wallis rank sum test was used. All four comparisons not highlighted were significant ($P < 0.0001$). (**F**) Prenylation status (p46-negative or p46-positive) determined by the presence or absence of p46 transcript from (D) in mild and severe COVID-19. (**G**) Tabulated ORs and 95% CIs of the data presented in (D) and (F). (**H**) SARS-CoV-2 infectious titer on AAT cells expressing the OAS1 isoforms p46 or p46 and p42. Isoform expression level (Western blot) is also shown.

structures reminiscent of the endoplasmic reticulum (Fig. 4I), whereas nonprenylated p42 and p46 C397A were diffusely distributed (Fig. 4I). To determine whether prenylated OAS1 localized to SARS-CoV-2 replicative organelles, we co-stained infected cells with the viral nsp5 (*8*). However, the p46 block was sufficiently strong to prevent formation of nsp5-positive replicative structures, which were only visible in clusters of cells expressing low levels of p46 (Fig. 4I). To overcome this, we imaged infected OAS1-expressing cells in which RNase L expression was disrupted using CRISPR-Cas9. Relieving the block to SARS-CoV-2 replication imposed by OAS1 facilitated the visualization of SARS-CoV-2 replicative structures (Fig. 4I and fig. S4B). Although OAS1 expression was enriched in close proximity to nsp5, these proteins did not appear to colocalize. We thus examined the colocalization of OAS1 and the corresponding dsRNA

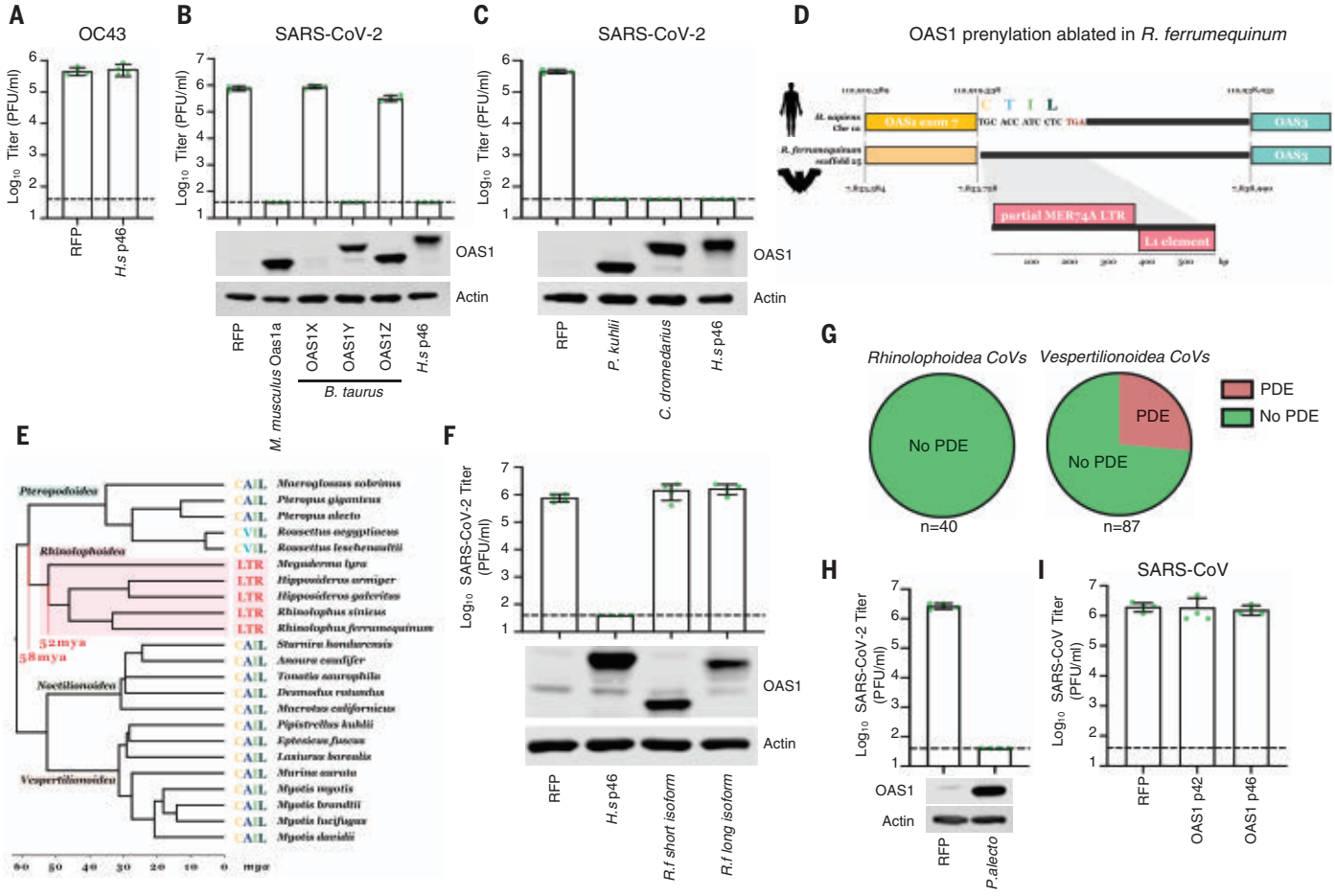


Fig. 6. Retrotransposition at the OAS1 locus has ablated the CAAX-box prenylation signal in *Rhinolophoidea*. (**A**) Infectious titers of OC43 (PFU/ml) were determined on AAT cells modified to express OAS1 from humans (*H.s* p46). (**B**) Infectious titers of SARS-CoV-2 CVR-GLA-1 (*8*) (PFU/ml) were determined on AAT cells modified to express Oas1a from mouse (*M. musculus*), OAS proteins from cows (*B. taurus*), and human p46 (*H.s* p46). OAS1 expression was monitored by Western blotting (lower panels). (**C**) Infectious titers of SARS-CoV-2 (PFU/ml) were determined on AAT cells modified to express OAS1 from *Pipistrellus* bats (*P. khulii*), dromedary camels (*C. dromedarius*), and human p46 (*H.s* p46). OAS1 expression was monitored by Western blotting. (**D**) Schematic of genome synteny between the human OAS1 exon 7 locus (yellow) and the *R. ferrumequinum* genome. The exact syntenic sequence coordinates are annotated for the start of OAS1 exon 7, the start of the CAAX box encoding sequence, and the start of the upstream gene locus, OAS3 (blue). Transposable element hits on the 580-bp nonsyntenic region in the *R. ferrumequinum* genome are shown in the enlarged inset. Noncoding regions are shown in black. Note that the schematic is not to scale. (**E**) Dated phylogeny (retrieved from TimeTree; www.timetree.org) (*101*) of bat species with a confirmed LTR insertion in the OAS1 locus or a CAAX box–encoding sequence present in the same scaffold as their OAS1 locus. Clades are labeled by superfamily, species names, and CAAX sequence (or LTR) are annotated next to the tree tips. The approximate time period during which the LTR insertion took place is annotated in red. (**F**) Infectious titers of SARS-CoV-2 CVR-GLA-1 (PFU/ml) were determined on AAT cells modified to express OAS1 from humans (*H.s* p46) and horseshoe bats (*R.f*) using both NCBI and Ensembl database entries. OAS1 expression was monitored by Western blotting. (**G**) Pie charts of CoVs from *Rhinolophoidea* and *Vespertilionoidea* binned according to whether they are known or predicted to encode a PDE OAS antagonist. (**H**) Infectious titers of SARS-CoV-2 (PFU/ml) were determined on AAT cells modified to express OAS1 from the black fruit bat (*P. alecto*). OAS1 expression was monitored by Western blotting. (**I**) Infectious titers of SARS-CoV (PFU/ml) were determined on AAT cells modified to express human OAS1 p42 or p46.

pathogen-associated molecular pattern. In the absence of RNase L, SARS-CoV-2 dsRNA colocalized with prenylated OAS1 (Fig. 4, J and K). By contrast, dsRNA detection overlapped poorly with nonprenylated p42 (Fig. 4, J and K, and fig. S4C). Considered alongside the iCLIP experiments (Fig. 3, I to K, and fig. S3), these data indicate that prenylation targets OAS1 to sites rich in viral dsRNA, which are probably the SARS-CoV-2 replicative organelles. Once in the right place, OAS1 binds to dsRNA structures in the SARS-CoV-2 5′-UTR and initiates a potent block to SARS-CoV-2 replication.

Prenylated OAS1 is associated with less severe COVID-19

The realization that prenylation is essential for OAS1-mediated sensing of SARS-CoV-2 allowed us to examine the transcriptome of infected patients and investigate whether there is a link between the expression of prenylated OAS1 and SARS-CoV-2 disease progression. The four most common p46 variants (Fig. 5A) all conferred protection against SARS-CoV-2 infection when exogenously expressed (Fig. 5B). Because each p46 variant had antiviral activity (i.e., was not confined to a single haplotype), we examined the p46 splice junction directly (Fig. 4A) to assess whether expression of p46-encoding mRNA, as opposed to the presence or absence of SNPs (*1*), significantly influenced the severity of COVID-19. The frequency of the Rs10774671 G SNP that governs the expression of p46 varies between ~11% (population: Peruvian, in Lima) and ~70% (population:

Esan, in Nigeria) (Fig. 5C) and could have a major influence on the susceptibility of different populations to severe COVID-19.

Although the p46 transcript encodes 54 C-terminal amino acids that are not part of the p42 protein (Fig. 4B), individuals homozygous for A at Rs10774671 (AA) can form splice junctions 1 nucleotide downstream of the p46 splice junction. We therefore confirmed that we could reliably detect the absence of p46 in RNA-sequencing (RNA-seq) data derived from cells with an AA genotype. We used JunctionSeq to examine all OAS1 transcript junctions (annotated on Ensembl) in these AA cells in the presence or absence of IFN treatment (fig. S5, A to C). Accordingly, we were unable to detect the specific junction (J080) between exons 5 and 7 that specifies the expression of prenylated OAS1 in these cells (fig. S5, A to C).

We therefore applied this method to RNA-seq data from 499 hospitalized UK COVID-19 patients with known disease outcomes (ISARIC4C) (fig. S6, A to C). We defined severe outcomes as intensive care unit (ICU) admission and/or death, and mild outcome patients as those who were hospitalized but not admitted to an ICU. All patients expressed detectable OAS1, but 42.5% of individuals (212/499) did not express p46. The absence of prenylated OAS1 was associated with more severe disease (Fig. 5, D to G). Specifically, the median transcript abundance of p46 was >100-fold lower in the severe COVID-19 group (Fig. 5D). This difference was entirely driven by the overrepresentation of patients in the severe COVID-19 group who did not express any prenylated OAS1. p46 mRNA levels were almost identical in individuals who expressed prenylated OAS1 regardless of whether they experienced mild or severe COVID-19 (Fig. 5D). Similarly, increased p42 expression was also associated with more severe COVID-19 (Fig. 5E). However, this association appeared to be a surrogate measurement of p46 expression, because patients who did not express any prenylated OAS1 expressed substantially higher levels of p42 (Fig. 5E). Crucially, no difference in p42 expression was apparent once the ability to express p46 was considered (Fig. 5E). Patients lacking the p46 transcript were more frequently observed in the severe disease group (Fig. 5F) and were significantly more likely to experience severe disease (95/212, 44.8%) compared with those expressing p46 (98/287, 34.1%) [unadjusted odds ratio (OR) =1.57, 95% confidence interval (CI) = 1.09 to 2.25; after adjustment for age, sex, and ethnicity and exclusion of 30 cases with missing data, OR = 1.58, 95% CI = 1.08 to 2.30] (Fig. 5G). Death was also more frequent in these patients (34/212, 16% versus 34/287, 11.8%) with the effect size similar to that for disease severity, but these differences were not statistically significant (unadjusted OR = 1.42, 95% CI = 0.85 to 2.37). Because higher p42 expression was associated with increased disease severity (Fig. 5E), we investigated whether p42 influenced the ability of p46 to inhibit SARS-CoV-2. Consistent with the lack of inhibition observed in Fig. 4C, p42 did not substantially blunt the ability of p46 to initiate a block to SARS-CoV-2 (Fig. 5H). Again, this is consistent with the association of high p42 expression with severe COVID-19 being mechanistically underpinned by an absence of p46.

Previous studies have identified an OAS1 haplotype that was inherited from Neanderthals and was associated with reduced susceptibility to SARS-CoV-2 infection and protection from severe COVID-19 (*22*, *24*). Our evaluation of OAS1 antiviral activity in vitro, combined with our analysis of OAS1 transcripts in patient cohorts, indicates that the protective Neanderthal OAS1 haplotype (*22*, *24*) likely prevents severe disease by specifying the expression of prenylated OAS1, which directs dsRNA sensing to the sites of SARS-CoV-2 replication. Indeed, by combining multiple studies, Huffman *et al.* also concluded that the Rs10774671 SNP is responsible for the protection conferred by the Neanderthal haplotype (*59*).

An ancient retrotransposition event ablated OAS1 prenylation in horseshoe bats

Every species has a specific repertoire of genome-encoded antiviral defenses (*26*). Because the differential splicing of p46 and p42 isoforms is poorly characterized beyond primates, it was previously difficult to investigate the protection conferred by p46 in nonhuman species. The realization that prenylation can be essential for antiviral activity allowed us to investigate this aspect of OAS1 biology beyond humans. For example, many coronaviruses encode phosphodiesterases (PDEs) that degrade 2-5A and antagonize the OAS system (*60*). The human betacoronavirus OC43 encodes such a phosphodiesterase (NS2) (*61*) and, accordingly, we found that prenylated OAS1 did not inhibit OC43 (Fig. 6A). Similarly, the Middle East respiratory syndrome-related CoV (MERS-CoV) also encodes a PDE (NS4b) capable of antagonizing the OAS system (*62*). We hypothesized that the reservoir species of OC43 and MERS-CoV encode OAS proteins that can initiate a block to CoV replication. OC43 likely originated in a murine host (*63*) and entered human populations through a cross-species transmission from cows (*64*). Examination of mouse and cow OAS1 sequences identified eight murine paralogs, three of which have CAAX boxes (Oas1a, Oas1f, and Oas1g), and three bovine paralogs, one of which has a CAAX box (OAS1Y). Consistent with Fig. 4, prenylated murine OAS1a and prenylated bovine OAS1Y both conferred potent anti–SARS-CoV-2 activity, whereas nonprenylated bovine OAS1Z did not (Fig. 6B, note that we were unable to confirm efficient expression of bovine OAS1X using polyclonal antibody raised to human OAS1).

Close relatives of MERS-CoV have been identified in bats (such as *Pipistrellus kuhlii*) (*65*) and MERS-CoV entered human populations after transmission from dromedary camels (*Camelus dromedarius*) (*66*, *67*). Accordingly, OAS1 from *P. kuhlii* and *C. dromedarius* have CAAX boxes and both instigated potent blocks to SARS-CoV-2 (Fig. 6C). Notably, the *P. kuhlii* C terminus is shorter than human p42, reinforcing the notion that most of the extended C terminus of p46 is not necessary for antiviral activity (fig. S7A). Crucially, this means that all of the species believed to harbor either OC43 or MERS-CoV en route to emergence in humans express prenylated OAS1 proteins that could credibly have selected for the maintenance of PDE expression in these viruses.

The extreme sensitivity of SARS-CoV-2 to prenylated OAS1 also led us to investigate whether horseshoe bats, the likely source of SARS-CoV-2, have a prenylated OAS1 defense. There is a paucity of mRNA sequence data available for horseshoe bats, and we were unable to find OAS1 database entries from the likely bat hosts of the precursors of SARS-CoV-2 [*Rhinolophus affinis* or *Rhinolophus pusillus* (*68*)] and SARS-CoV (*Rhinolophus sinicus*). Thus, we analyzed OAS1 transcripts from the greater horseshoe bat, *Rhinolophus ferrumequinum*. We were unable to find a transcript or exon encoding a prenylated OAS1 in *R. ferrumequinum*. Indeed, all Ensembl and National Center for Biotechnology Information (NCBI) database entries specified nonprenylated proteins. When we examined the genomic region where the prenylation signal should reside (on the basis of synteny and homologous flanking sequences), retrotransposition of a long terminal repeat (LTR) sequence was evident, and this ablated the CAAX-box motif, preventing the expression of prenylated anti–CoV OAS1 in these bats (Fig. 6D). We searched for this insertion in 44 available bat genome sequences and identified the same insertion only in members of the Rhinolophoidea superfamily (including *Rhinolophus*, *Hipposideros*, and *Megaderma* species), indicating that this ancient retrotransposition insertion occurred ~58 to 52 million years ago within this bat superfamily. By contrast, we could detect CAAX-box encoding syntenic sequences in members of all other bat taxa (Fig. 6E).

Because of the absence of prenylated OAS1 in Rhinolophoidea, we predicted that OAS1 from *R. ferrumequinum* would be inactive against SARS-CoV-2. Accordingly, the best supported OAS1 isoforms from the greater horseshoe bat (fig. S7B) did not inhibit SARS-CoV-2 (Fig. 6F). This means that membrane-associated, prenylated, p46-like OAS1 dsRNA sensing has been ablated in the presumed bat

reservoirs of SARS-CoV and SARS-CoV-2, and this may contribute to horseshoe bats being such prolific reservoir hosts of Sarbecoviruses. The absence of pressure to evade prenylated OAS1 (in horseshoe bats) may have left SARS-CoV-2 particularly sensitive to this defense (when subsequently encountered in human populations).

Considering the lack of prenylated OAS1 in Rhinolophoidea and the ability of coronavirus PDEs to antagonize this pattern recognition pathway, we investigated whether PDE-encoding coronaviruses infect horseshoe bats. Given the variability in coronavirus-encoded PDEs [NS4b in Merbecoviruses and NS2 in Embecoviruses (*61*, *62*)], we developed a custom HMM protein profile using NS4b, NS2, the mammalian PDE AKAP7 (*69*), and rotavirus A VP3 (*70*). We screened for PDEs through all putative open reading frames (ORFs) of all published Coronaviridae genomes. This method should capture previously unannotated or undescribed PDEs. Although the available sequence data set is likely biased by sampling, we could not identify PDEs in any known coronaviruses from Rhinolophoidea. In fact, all of the bat coronaviruses identified as encoding PDEs were from bats in the Vespertilionoidea superfamily (in which prenylated OAS1 is intact) (Fig. 6G). Although there is an appreciable absence of PDEs in the CoVs that circulate in horseshoe bats, an absence of PDEs does not necessarily imply an absence of anti–CoV OAS proteins in the relevant host. Many potential strategies exist to evade or antagonize the OAS system (*46*), and we also did not identify PDEs in CoVs sampled from Pteropodoidea. To confirm that prenylated OAS1 from megabats have anti-CoV potential, we observed that OAS1 from *Pteropus alecto* could instigate a potent block to SARS-CoV-2 (Fig. 6H).

Before the COVID-19 pandemic, SARS-CoV emerged in humans in 2003. Close relatives of SARS-CoV circulate in *R. sinicus (71)*, and SARS-CoV was transmitted to humans through an intermediate species widely believed to be civets (*72*). When we considered the ability of human OAS1 to inhibit an isolate of SARS-CoV, this virus was completely resistant to human p46 and p42 (Fig. 6I). Because the retrotransposition event that ablates the prenylation sequence was confirmed in *R. sinicus* (Fig. 6E), we speculate that SARS-CoV may have acquired the ability to evade or antagonize OAS1 during circulation in an intermediate species or in human populations.

Discussion

Viruses tend to rapidly adapt to new host species, and even SARS-CoV-2, a "generalist" virus (*73*), has likely adapted to replicate in the animal reservoir(s) in which it circulated before emergence in humans. Cross-species transmission exposed SARS-CoV-2 to a new repertoire of human antiviral defenses, some of which the virus may not have encountered before. Prenylated OAS1 may be an example of such a defense, with prenylated OAS1 being targeted to endomembranous structures where it initiates potent anti–SARS-CoV-2 activity in vitro.

Hospitalized COVID-19 patients lacking the p46 transcript had worse clinical outcomes than those who expressed prenylated OAS1. Severe disease was significantly more frequent, with ICU admission or death being ~1.6 times more likely in these patients. The increased odds of death were similarly raised among patients lacking p46. However, the number of deaths in this cohort was relatively small, and a larger study would be needed to provide enough power to determine whether a lack of p46 transcript is also associated with increased mortality. In addition, we could not detect an association between disease severity and p46 transcript abundance in individuals who expressed p46. This could reflect the fact that expression level is less important for an enzyme because catalysis greatly amplifies pattern recognition. It should also be considered that p46 expression in whole blood may not recapitulate expression differences at the sites of viral replication or important differences present early in infection (before hospitalization). Along these lines, OAS1 levels have been linked to the severity of COVID-19 in other research (*22*). Although prenylated OAS1 tips the balance in favor of the host in a significant number of people, OAS1 is just one component of a "successful" immune response, and it is likely that multiple ISGs (alongside many other factors) influence the outcome of SARS-CoV-2 infection.

Considering the apparent lack of antiviral activity of OAS1 p42, it is notable that the p42-encoding alleles predominate and are more common in all human populations (apart from people of African descent). The geographical variation in the frequency of alleles encoding prenylated OAS1, such as the high frequency observed in some African populations, could potentially influence the spread and severity of COVID-19. It is currently difficult to reconcile a potential role for p46 in reducing the impact of COVID-19 in some African countries (*59*, *74*) with the increased disease severity observed in people of African descent (*75*). None of our observations refute the possibility that p42 enhances the anti–SARS-CoV-2 activity of p46. It is possible that p42 may be more beneficial than p46 in some contexts, potentially targeting viruses that do not use replicative organelles. Alternatively, the p46 variant may have been selected against, possibly because it is deleterious in the absence of specific viral infections (*56*, *76*). In support of this notion, the catalytic activity of OAS1 has been lost entirely in some species (*56*). Because p46 may reduce the susceptibility to SARS-CoV-2 infection (*22*), the billions of individuals who are unable to express prenylated OAS1 may make humans relatively vulnerable to the direct cross-species transmission of Sarbecoviruses from horseshoe bats.

There is currently great interest in identifying the biological characteristics of bats that might predispose them to be reservoirs of circulating viruses (*77*), and much work has focused on innate immunity (*78*). It is important to be cautious when generalizing about bats because each species has specific innate immune features (*26*). Bats are an extraordinarily diverse order (>1400 species) (*79*), and individual bat species may not be any more likely to act as viral reservoirs than other species (*80*). Nevertheless, it is notable that horseshoe bats not only lack prenylated OAS1-mediated dsRNA sensing but also have a reduced ability to sense cytosolic DNA through stimulator of interferon genes (STING) (*81*). It is tempting to speculate that multiple defects in pathogen recognition may make horseshoe bats particularly good virus reservoirs. However, this requires resolution regarding why innate immune defects might promote tolerance in bats while promoting pathogenesis in humans.

The endomembrane targeting of prenylated OAS1 enables the potential sensing of a diverse spectrum of viruses. For example, multiple viruses that use replicative organelles, including hepatitis C virus (*82*), Alphaarterivirus equid (EAV) (*83*), and Betaarterivirus suid 1 (PRRSV) (*84*), are inhibited by OAS1. However, identification of the sites in viral and host RNAs that were bound by OAS1 underscored how selective a sensor OAS1 is. Few host RNAs were bound by OAS1 and despite replicating through a dsRNA intermediate, most sites in the SARS-CoV-2 genome escaped detection by OAS1. It appears that multiple layers of antiviral specificity exist for prenylated OAS1, and OAS1 likely only recognizes a specific subset of dsRNA sequences that occur close to endomembranes. The considerable target specificity of OAS1 is likely necessary because OAS1 senses relatively short stretches (~18 bp) of dsRNA (*13*), and a less discerning sensor would be inappropriately activated by cellular RNAs.

The ability of SARS-CoV to escape inhibition by human p46 highlights that whereas many viruses might conceivably be targeted, evasion and antagonism strategies (*46*) mean that sensitivity for every virus must be considered on a case-by-case basis. Consistent with this, although SARS-CoV diverges from SARS-CoV-2 at several positions within the 5′-UTR, it is currently unclear whether this is the strategy that enables SARS-CoV to escape from OAS1-mediated sensing.

Prenylated OAS1 has contributed to the prevention of severe COVID-19 in a substantial

fraction of infected individuals (likely measured in "hundreds of thousands" in the UK alone). With the continued emergence of SARS-CoV-2 variants of concern, it will be important to remain vigilant. The chance that recombinant acquisition of a PDE gene [from a coinfecting virus or host gene (*69*, *85*)] or repeated selection against SL1 and SL2 in the 5′-UTR could enable escape from this OAS1 defense and increase the pathogenicity of SARS-CoV-2. This reinforces the need to pay close attention to the phenotypic properties of emerging SARS-CoV-2 variants.

Materials and Methods

Cell lines, plasmids, and viruses

All cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 9% fetal calf serum (FCS) and 10 μg/ml gentamicin unless otherwise stated. A549-ACE2-TMPRSS2 ("AAT") and VeroE6-ACE2-TMPRSS2 ("VAT") cells have been described previously (*8*). Human embryonic kidney (HEK) 293T cells were propagated from laboratory stocks, Vero E6 cells were a generous gift of M. Bouloy, and A549-Npro cells were a kind gift of R. E. Randall. HT1080 cells were a kind gift of S. Neil and were modified to overexpress human ACE2 and TMPRSS2 (referred to as "HAT" cells) and were transduced as described previously (*8*). Calu-3 cells were a generous gift from P. J. Lehner and were maintained in MEM supplemented with 10% FCS, 2 mM glutamine, 2 mM sodium pyruvate, and 100 μM nonessential amino acids.

The SARS-CoV-2 viruses CVR-GLA-1, England/02/2020 and SARS-CoV-2-ZsGreen have been described previously (*8*). SynSARS-CoV-2-eGFP was a kind gift from V. Thiel (*11*). SARS-CoV-2 lineage B.1.1.7 isolate "212" was isolated from a clinical sample (kind gift of W. Barclay). VSV was a kind gift of M. Stanifer (*86*). Influenza A viruses A/Puerto Rico/8/1934 (H1N1) and A/Mallard/Netherlands/10-Cam/1999(H1N1) were rescued from reverse genetics systems (a kind gift from R. Fouchier, and L. Tiley, respectively) as described previously (*87*, *88*). Human respirovirus 3 with GFP (PIV3-GFP) was purchased from ViraTree. Human RSV-GFP was a kind gift from P. Collins (*89*). Cardiovirus A (EMCV) was a kind gift from C. Bamford. Betacoronavirus OC43 (ATCC VR1558) was purchased from ATCC and propagated on VAT cells (*8*). SARS-CoV virus isolate (HKU39849, GenBank: AY278491.2) was a kind gift of M. Peiris and supplied by B. Haagmans.

Retroviral vectors and cell modification

The lentiviral vector pSCRPSY (KT368137.1) has been previously described (*5*). pLV-EF1a-IRES-Puro (Addgene plasmid #85132) or pLV-EF1a-IRES-Blast (Addgene plasmid #85133) were modified by polymerase chain reaction (PCR) amplifying the TagRFP ORF (using pSCRPSY as template) flanked by directional *Sfi*I sites, which were further flanked by *Bam*HI and *Eco*RI restriction sites (forward oligo: 5′-CTC TCG GAT CCG GCC GAG AGG GCC ATG AGC GAG CTG ATT AAG-3′ and reverse oligo: 5′-CTC TCG AAT TCG GCC AGA GAG GCC TCA CTT GTG CCC CAG-3′), and the *Bam*HI-*Eco*RI fragment was subcloned into the vectors to create the modified pLV-EF1a-IRES-Puro-*Sfi*I-TagRFP or pLV-EF1a-IRES-Blast-*Sfi*I-TagRFP constructs. The cDNA corresponding to the ORFs of the following OAS genes (GenBank accession number): OAS1p46 (NM_016816), human OAS3 (NM_006187), mouse Oas1A (NM_145211), bovine OAS1X (NM_178108), bovine OAS1Y (NM_001040606), bovine OAS1Z (AY650038), *P. kuhlii* (XM_036409709.1), *C. dromedarius* (XM_031443284), *R. ferrumequinum* OAS1 (short isoform: XM_033097132 / long isoform: ENSRFET00010016745), and *P. alecto* (NM_001290162) were synthesized as gene blocks with flanking *Sfi*I sites (IDT DNA), and the *Sfi*I fragment was subcloned into the modified pLV-EF1a-IRES-Puro-*Sfi*I plasmid. To generate the human OAS1p42 sequence (in accordance with GenBank accession NM_002534), OAS1p46-C397A, and OAS1p42-CTIL sequences, the pLV-*Sfi*I-OAS1p46 lentiviral vector plasmid was modified by overlap extension PCR (using primer pair 5′- CTC TCT GGC CGA GAG GGC CAT GAT GGA TCT CAG AAA TAC CCC AG-3′ and 5′- TCT CTC GGC CAG AGA GGC CTC AAG CTT CAT GGA GAG GGG CAG GGA TGA ATG GCA GGG AGG AAG CAG GAG GTC TCA CCA GCA GAA TCC AGG AGC TCA CTG GG-3′ for OAS1p42, primer pair 5′- CTC TCT GGC CGA GAG GGC CAT GAT GGA TCT CAG AAA TAC CCC AG -3′ and 5′- TCT CTC GGC CAG AGA GGC CTC AGA GGA TGG TGG CGG TCC AGT CCT CTT CTG CCT GTG GG -3′ for OAS1p46-C397A, and primer pair 5′- CTC TCT GGC CGA GAG GGC CAT GAT GGA TCT CAG AAA TAC CCC AG -3′ and 5′- TCT CTC GGC CAG AGA GGC CTC AGA GGA TGG TGC AAG CTT CAT GGA GAG GGG CAG GGA TGA ATG GCA GGG AGG AAG CAG GAG GTC TCA CCA GCA GAA TCC AGG AGC TC ACT GGG -3′ for OAS1p42-CTIL) and the respective *Sfi*I fragments were subcloned in place of OAS1p46 in the pLV lentiviral vector plasmids described above. To generate the human OAS1 p46 Δ12 aa and Δ32 aa C-terminal truncations, the pLV-*Sfi*I-OAS1p46 lentiviral vector plasmid was modified by overlap extension PCR (using primer pair 5′-CTC TCG GAT CCG GCC GAG AGG GCC- 3′ and 5′ -TCT CTC GGC CAG AGA GGC CTC ATC AGA GGA TGG TGC ACT GGA GTG TGC TGG G-3′ for OAS1 p46 Δ12 aa, using primer pair 5′-CTC TCG GAT CCG GCC GAG AGG GCC- 3′ and 5′ -TCT CTC GGC CAG AGA GGC CTC AGA GGA TGG TGC ATT TCT GAT ACC TCC TGG GAT CGT- 3′ for OAS1 p46 Δ32 aa) and the respective *Sfi*I fragments were subcloned into the pLV lentiviral vector plasmids described above. The four most frequent OAS1 p46 protein haplotypes as shown on Ensembl were synthesized with flanking *Sfi*I sites and subcloned into pLV-EF1a-IRES-Puro-*Sfi*I-TagRFP by Genewiz. Lentiviral vectors were produced by transfecting HEK 293T cells as described previously (*86*), 0.45-μm pore size filtered supernatant was used to transduce AAT cells (*8*), and transduced cells were selected using 2 μg/ml puromycin or 5 μg/ml blasticidin.

Gene editing by CRISPR-Cas9 was achieved using the lentiCRISPRv2-BlastR or lentiCRISPRv2-PuroR one vector system following the established protocols from the Zhang laboratory. CRISPR guides were designed using the CHOPCHOP online tool (https://chopchop.cbu.uib.no). Seven guides and one nontargeting guide per target were subcloned into the one vector system between the *Bsm*BI sites using annealed oligonucleotides with directional, compatible *Bsm*BI overhangs and tested for their efficacy to ablate RNase L, OAS3 or OAS1 expression, respectively. The following target sequences for the guides were used: nontargeting control guide ("NTC": 5′-GTG ACG TAC CGC TGG AGG TA-3′), RNase L guides (guide 1: 5′-GCC GAG TTG CTG TGC AAA CG-3′, guide 2: 5′- TTA TCC TCG CAG CGA TTG CG-3′, guide 3: 5′-CTA TAG GAC GCT TCG GAA TG -3′, guide 4: 5′-TAT AGG ACG CTT CGG AAT GT-3′, guide 5: 5′-TAG TCA TCT TCA GCC GCT AT-3′, guide 6: 5′-TTT ATC CTC GCA GCG ATT GC-3′, and guide 7: 5′-GCA ATC GCT GCG AGG ATA AA -3′), OAS3 guides (guide 1: 5′-CAT CAA GGA TCT CTG CGC GG- 3′, guide 2: 5′ -TCA AGG ATC TCT GCG CGG CG- 3′, guide 3: 5′ -CTT GGG TTT GAC GCC GGA GC- 3′, guide 4: 5′ -CGT TCC AGG TGG GAT CAG CG- 3′, guide 5: 5′ -CAA GAT CTA CGG ATG TCA GG- 3′, guide 6: 5′ -AAT TCC AGG GCA TAG ACC GG- 3′, guide 7: 5′ -GAC AGT TTT CAG CAC CCG CG- 3′), OAS1 guides (guide 1: 5′ -TCA TCC GCC TAG TCA AGC AC- 3′, guide 2: 5′ -CGG TCT ATG CTT GGG AGC GA- 3′, guide 3: 5′ -TGC ATG CGG AAA CAC GTG TC- 3′, guide 4: 5′ -AAG TTT CCT GTA GGG TCC GC- 3′, guide 5: 5′ -GTA CGA AGC TGA GCG CAC GG- 3′, guide 6: 5′ -AAT CTA TGT CAA GCT CAT CG- 3′, guide 7: 5′ -CGA ACA GGT CAG TTG ACT GG- 3′). RNase L and OAS3 guides were transduced into AAT-OAS1-p46 cells and subsequently selected and cultured in medium additionally supplemented with 5 μg/ml blasticidin (Melford Laboratories). OAS1 guides were transduced into HAT cells, selected, and cultured in medium additionally supplemented with 2 μg/ml puromycin (Melford Laboratories).

Arrayed ISG expression screening

The ISG overexpression libraries and flow cytometry–based screening have been described

previously (*5*, *6*, *86*, *90*). Briefly, two lentiviral vector ISG libraries consisting of 539 human and 444 macaque ISGs were used to transduce A549-Npro-ACE2 cells (1.25×10^4 cells/well in a 96-well plate seeded the day before) in the presence of polybrene for 48 hours (aiming for an average of >90% transduction), allowing ISG expression from an early HIV-1 mRNA and TagRFP expression from an unspliced late HIV-1 mRNA, the latter used as a marker for transduction. Transduced cells were then infected with synSARS-CoV-2-eGFP (*11*) in the presence of DMEM supplemented with 2% FCS. At 14 or 40 hours after infection, cells were trypsinized and fixed in 4% formaldehyde. The percentage of transduced cells (TagRFP-positive) and SARS-CoV-2 infected cells (GFP-positive) were determined by flow cytometry using a Guava EasyCyte flow cytometer (Millipore).

A549-ISRE::GFP cells [gift from R. E. Randall (*91*)] were transduced in the presence of polybrene with the "miniscreen" library of selected ISGs from the SARS-CoV-2 screening. Then, 96 hours after transduction, the supernatant was harvested to measure the toxicity of the expressed ISGs using the CytoTox-Glow kit (Promega), and cells were fixed in 4% formaldehyde to measure ISRE induction (GFP-positive cells) as a surrogate for IFN induction.

The screens in Fig. 3A followed the general scheme outlined above. Target cells were seeded in 96-well plates (0.1×10^5 to 0.6×10^5 cells/well depending on the cell line) either immediately before transduction (suspension cells) or the day before transduction (adherent cells). Cells were transduced with the SCRPSY ISG library as described above. At 48 hours after transduction, the transduced cells were split 1:2 (suspension cells) before challenge with a reporter-encoding virus [using a dose lower than multiplicity of infection (MOI) 1 to achieve 20-50% infection at the time of fixation]. The panel of viruses shown in Fig. 3A are as follows: AdV, Human mastadenovirus C (Adenovirus 5); PRV, Suid herpesvirus 1 (Pseudorabies); HSV-1, Human herpesvirus 1; BoHV-1, Bovine herpesvirus 1; RVFV, Rift Valley fever phlebovirus; SFTSV, Dabie bandavius (Severe fever with thrombocytopenia syndrome virus); BUNV, Bunyamwera orthobunyavirus; BTV, Bluetongue virus; Rotavirus, Simian Rotavirus A/SA11; CHPV, Chandipura vesiculovirus; VSV, Indiana vesiculovirus; SFV, Semliki Forest virus; CHIKV, Chikungunya virus; HIV-1, Human immunodeficiency virus 1; IAV PR8, A/Puerto Rico/8/1934 (H1N1); IAV Cal04, A/California/04-061-MA/2009 (H1N1); IAV Mallard, A/Mallard/Netherlands/10-Cam/1999 (H1N1); MeV, Measles Ed-Zag vac; SeV, Murine respirovirus (Sendai virus); PIV5, Mammalian orthorubulavirus 5 (Parainfluenza virus 5 or simian virus 5); hMPV, human Metapneumovirus; hRSV, human orthopneumovirus (human respiratory syncytial virus); bRSV, Bovine orthopneumovirus (bovine respiratory syncytial virus); PIV3, human respirovirus 3 (Parainfluenza virus 3); and ZIKA, Zika virus. Further details regarding these viruses, cells, and time points used for screens presented in Fig. 3A are available from Enlighten (https://doi.org/10.5525/gla.researchdata.1178). ISG screens were single biological experiments that were followed up with confirmatory "miniscreens" executed in two biological replicates with a typical result being presented.

Virus infections and titrations

SARS-CoV-2 infection assays, plaque assays, and well-clearance assays have been described previously (*8*). Briefly, well-clearance assays quantify transmitted light (Celigo, Nexcelom) through imaging of a stained cell monolayer. CPE-induced clearance of the monolayer transmits more light compared with uninfected or protected monolayers. AAT cells were seeded at 1.25×10^4 cells/well and HAT cells at 2×10^4 cells/well and incubated overnight. The following day, cells were pretreated or not with IFN for 2 hours before infection with doses specified in the text and/or figure. Virus inputs were either normalized using plaque assays on VeroE6 cells to 500 plaque-forming units (PFUs) per well or virus input was titrated with threefold dilutions adding 50 μl/well.

For plaque assays, AAT derivative cells and VeroE6 cells were seeded in 12-well plates at 3×10^5 cells/well and incubated overnight. The following day, cells were inoculated with 250 μl of 10-fold logarithmic dilutions of virus stocks prepared in serum-free DMEM. After 1 hour of virus adsorption at 37°C, the wells were overlaid with 0.6% Avicel in MEM. After 3 days of incubation, plates were submerged in 8% formaldehyde, washed in phosphate-buffered saline (PBS), and stained with Coomassie blue for plaque visualization. Plaque assays with SARS-CoV, VSV, EMCV, and OC43 were performed under identical conditions to SARS-CoV-2, incubating for 2 days (SARS-CoV, VSV, and EMCV) or 5 days (OC43). IAV immunostaining of foci was achieved using the mouse anti–influenza A virus nucleoprotein monoclonal antibody clone AA5H (BioRad, MCA400) and visualized with goat anti-mouse IgG (H+L)–horseradish peroxidase conjugate (BioRad, 1721011) and TrueBlue Peroxidase Substrate (KPL, 5510-0030) following standard plaque assay protocols described previously (*88*).

For titration of GFP-encoding viruses (PIV3-GFP and RSV-GFP), 96-well plates of AAT derivative cells were seeded with 2×10^4 cells/well the day before. For titration of SARS-CoV-2-ZsGreen on Calu-3 derivative cells, 5×10^4 cells/well of a 96-well plate were seeded. The next day cells were infected with serial dilutions of virus for 24 hours (PIV3/RSV) or 40 hours (SARS-CoV-2). After incubation, the cells were trypsinized and the percentage of GFP-positive cells was measured by flow cytometry using a Guava EasyCyte cytometer (Millipore). All virus experiments represent a typical result of at least two biological repeats with four technical replicates (plaque assay) or three technical replicates (well-clearance and titration assays).

Western blot analyses

For preparation of cell lysates, cells were seeded in six-well plates with 1×10^6 cells/well the day before harvest. Cells were washed once with PBS, harvested in SDS sample buffer [12.5% glycerol, 175 mM Tris-HCl (pH 8.5), 2.5% SDS, 70 mM 2-mercaptoethanol, and 0.5% bromophenol blue] and then heated for 10 min at 70°C and sonicated. Proteins were subsequently separated on NuPage 4 to 12% Bis-Tris polyacrylamide gels and transferred onto nitrocellulose membranes. Blots were probed with either antibodies raised against actin (mouse JLA20 hybridoma; courtesy of the Developmental Studies Hybridoma Bank, University of Iowa), OAS1 (rabbit polyclonal 14955-1-AP, Proteintech), OAS2 (rabbit polyclonal 19279-1-AP, Proteintech), OAS3 (rabbit polyclonal 21915-1-AP, Proteintech), or the rabbit anti-RNase L monoclonal antibody (Cell Signaling Technology, 27281). Thereafter, membranes were probed with species IgG-specific fluorescently labeled secondary antibodies goat anti-rabbit IgG (Thermo Fisher Scientific, SA5-10036) or goat anti-mouse IgG (Thermo Fisher Scientific, SA5-10176) and scanned using a LiCor Odyssey scanner.

Immunofluorescence

Subconfluent AAT derivative cells seeded on glass coverslips were infected with CVR-GLA-1 at an MOI of 0.5 for 24 hours. Cells were fixed in PBS/8% formaldehyde, permeabilized with PBS/0.2% TX-100, and blocked with PBS/1% BSA. Immunostaining was performed using a rabbit anti-OAS1 monoclonal antibody [clone D1W3A] (Cell Signaling Technology, 14498) and sheep anti-SARS-CoV-2-nsp5 antiserum [https://mrcppu-covid.bio, described in (*8*)] or mouse anti-dsRNA monoclonal antibody (J2, Nordic MUBio 10010500). Secondary antibody staining was performed with Alexa Fluor 488 goat anti-rabbit IgG and Alexa Fluor 594 donkey anti-sheep IgG both at 1:1000 (Invitrogen). Hoechst 33342 was included in the secondary antibody stain at 5 μg/ml. Coverslips were mounted on glass slides (VWR) using ProLong Gold antifade mountant (Thermo Fisher Scientific).

Maximum intensity projection images of cell monolayers were acquired with an Airyscan Fast detector fitted to a Zeiss LSM880 confocal microscope. The objective lens used was a Plan-Apochromat 63×/1.4 oil DIC M27 (Carl Zeiss) and gain, laser power, and pinhole were synchronized across images. Maximum

intensity projection images with pixel scaling of 0.04 μm × 0.04 μm each comprised a *Z*-stack of 10 individual slices with a total focal depth of 1.435 μm. The OAS1–Alexa Fluor 488 was excited at 488 nm and detected in the 495 to 550 nm range, nsp5–Alexa Fluor 594 was excited at 594 nm and detected with a long-pass 605 nm filter, and Hoechst was excited at 405 nm and detected in the 420 to 480 nm range. After acquisition, the contrast of images within each set were optimized using Zen software (Carl Zeiss) to equal degrees for the vector, p42, p42-CTIL, and p46 C397A samples, and the histogram maximum was increased independently in the p46 sample shown in Fig. 4I to prevent oversaturation in the green channel caused by strong perinuclear concentration. Images were acquired as 8-bit *.czi files and exported as 8-bit TIFF files.

For dsRNA and OAS1 colocalization analysis, single-slice images were acquired as above but with a gallium arsenide phosphide photomultiplier tube (GaAsP-PMT) detector. The pinhole size was adjusted to obtain an equal focal depth of 2 μm in each channel. Images in each channel were acquired sequentially: OAS1–Alexa Fluor 488 was excited at 488 nm and detected in the 493 to 608 nm range, dsRNA–Alexa Fluor 594 was excited at 594 nm and detected in the 599 to 735 nm range, and Hoechst was excited at 405 nm and detected in the 426 to 500 nm range. To quantify colocalization of dsRNA with OAS1, the colocalization tool in Zen 3.2 software (Blue version, Carl Zeiss) was used to generate a weighted colocalization coefficient for the dsRNA–Alexa Fluor 594 channel in infected cells. Each data point represents a distinct region of interest encompassing an individual cell. Presented confocal images are from one of at least two independent experiments performed on separate days.

In situ hybridization

Formalin-fixed and paraffin-embedded (FFPE) lung tissue of two patients with confirmed SARS-CoV-2 infection (C21-20: 79 years old, male; C19-20: 56 years old, male) were used. As control tissue, FFPE lung of a healthy, 62-year-old male donor was used (NBP2-30182, Novusbio, catalog no. 0028000B). Additionally, FFPE respiratory nasal epithelium (Amsbio, code: AMS-41022-NE) from three healthy donors (patient 3: specimen 100229, a 35-year-old female; patient 4: specimen 100235, a 31-year-old female; and patient 5: specimen 100233, a 37-year-old female) was used.

For the detection of gene-specific RNA by in situ hybridization, the RNAscope 2.5 HD Reagent Kit-RED (code: 322350, Advanced Cell Diagnostics) and the probes (code: NCOA7 1029911-C1, ZBTB42 1029921-C1, OAS1 1029931-C1, ANKFY1 1029941-C1, UNC93B1 1029951-C1, SCARB2 1029961-C1; Advanced Cell Diagnostics) were designed (gene bank no: NM_001199622, NM_001137601, NM_016816, NM_016376, NM_030930, NM_005506) and purchased. As positive and negative controls, a human ubiquitin and a plant probe were used, respectively (codes: 310041 and 310043, Advanced Cell Diagnostics). The protocol was followed according to the manufacturer's instructions. This work (ethics approval number 32077020.6.0000.0005) was approved in May 2020 by the National Committee in Ethics and Research, Brazil in COMISSÃO NACIONAL DE ÉTICA EM PESQUISA. Informed consent was obtained from all study participants.

OAS1 iCLIP2

iCLIP procedures were performed according to iCLIP2 (*35*), with additional modifications (*36*). In detail, 3.5×10^6 AAT-OAS1p46-RNase L CRISPR Guide 5 cells were seeded into 10-cm tissue culture plates. The following day, cells were either infected with SARS-CoV-2 CVR-GLA-1 at MOI 2 in a 6-ml volume of DMEM supplemented with 2% FCS or mock treated with medium only. After 1 hour of virus or mock adsorption, the medium was changed to 12 ml of fresh DMEM supplemented with 2% FCS and incubated for 24 hours. Cells were washed once in ice-cold PBS before ultraviolet irradiation at 254 nm with 150 mJ/cm^2. Cells were lysed in 1 ml of iCLIP lysis buffer per plate [50 mM Tris-Cl (pH 7.4), 100 mM NaCl, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate, 0.2 mM AEBSF] for 30 min at 4°C. The experiment was performed in three biological replicates per mock and SARS-CoV-2 infected sample, prepared on separate days with two different virus stocks. Per sample, four units of TurboDNase (Life Technologies, AM2238) were added and lysate precleared by centrifugation for 10 min at 16,000*g* at 4°C. A further four units of TurboDNase and 10 units of RNase I (Life Technologies, AM2294) per sample were added and incubated for 3 min at 37°C at 1100 rpm. Then, 200 U Ribolock (Life Technologies, EO0382) per sample was added and incubated for 3 min on ice. Lysates were precleared with 50 μl of Protein A/G Sepharose (Thermo Fisher Scientific, 20423) for 30 min at 4°C with rotation. Supernatants were split and incubated with either 7.5 mg of OAS1 rabbit monoclonal antibody (clone D1W3A, Cell Signaling Technology, 14498S) or 7.5 mg of rabbit isotype control (BD Biosciences, 550875) for 45 min at 4°C with rotation, followed by binding with further 50 μl Protein A/G Sepharose for 45 min at 4°C. Beads were washed twice with 1 ml of high-salt buffer [50 mM Tris-Cl (pH 7.4), 1 M NaCl, 1mM EDTA, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate, and 0.2 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF)], twice with 1 ml of medium-salt buffer [50 mM Tris-Cl (pH 7.4), 250 mM NaCl, 1 mM $MgCl_2$, 0.05% Igepal CA-630, and 0.2 mM AEBSF], and twice with 1 ml of PNK wash buffer [20 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, and 0.2% Tween-20]. RNA was dephosphorylated at 37°C for 40 min at 1100 rpm in PNK buffer [50 mM Tris-HCl (pH 6.5), 10 mM $MgCl_2$, and 1 mM dithiothreitol (DTT)] with 5 U PNK (New England Biolabs, M0201L), 0.25 U FastAP alkaline phosphatase (Thermo Fisher Scientific, EF0654), 0.5 U TurboDNase, and 20 U Ribolock. Beads were washed once with PNK wash buffer, twice with high-salt buffer and twice with PNK wash buffer. L3-IR-adapter were generated after irCLIP (*92*) with the sequence /5Phos/AGATCGGAAGAGCGGTTCAGAAAAAAAAAAAA/iAzideN/AAAAAAAAAAAA/3Bio/ ordered from IDT. The adapter was ligated using 10 U T4 RNA ligase (Fisher Scientific, 10669690), 20 U Ribolock, 4 U PNK, 18% PEG 8000 (Sigma-Aldrich, P1458-25ML), 5% dimethyl sulfoxide (DMSO) at 16°C 1100 rpm overnight in the dark. Beads were washed once with PNK wash buffer, twice with high-salt buffer, and twice with PNK wash buffer. Samples and inputs were denatured in 1× NuPage lithium dodecyl sulfate sample buffer (Thermo Fisher Scientific, NP0007) with 100 mM DTT at 70°C for 5 min and separated on a 4 to 20% Mini-PROTEAN TGX (BioRad, 4561093). Protein-RNA complexes were transferred onto an iBLOT2 nitrocellulose membrane (Fisher Scientific, IB23001) and visualized using a Licor Odyssey CLx. The regions corresponding to 42 to 130 kDa were excised and digested using 20 μl of Proteinase K (Roche, 3115828001) in 10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM EDTA, 0.2% SDS at 50°C for 60 min at 1100 rpm. RNA was purified with acidic (pH 6.6 to 6.9) phenol:chloroform:isoamyl alcohol (Sigma-Aldrich, P3803), followed by RNA cleanup using Zymo RNA Clean & Concentrator-5 (ZYMO Research, R1013). A size-matched input (SMI) control was extracted in parallel and adapter ligation performed. The SMI was treated with 5 U PNK and 0.5 U FastAP and 20 U Ribolock in PNK buffer for 40 min at 37°C at 1100 rpm, followed by MyONE silane beads (Life Technology, 37002D) purification. L3-IR-adapter ligation was performed with 15 U T4 RNA ligase in 1× ligase buffer with 1.5% DMSO, 16% PEG8000, and 0.1 μM L3-IR-adapter for 75 min at room temperature, followed by MyONE purification. SMI was treated with 25 U 5′ deadenylase (New England Biolabs, M0331S) and 15 U RecJf endonuclease (New England Biolabs, M0264S) and 20 U Ribolock in 1× New England Biolabs buffer 2 with 8% PEG8000 for 1 hour at 30°C, and then 30 min at 37°C at 1100 rpm, followed by a MyONE bead purification. RNA from SMI and iCLIP samples were reverse transcribed using Superscript IV (Invitrogen, 18090010). RNA was hydrolyzed by the addition of 1.25 μl of 1 M NaOH and incubation at 85°C for 15 min before neutralization with 1.25 μl of 1 M HCl. cDNA was

purified using MyONE silane beads (Life Technology, 37002D). Adapters specific to each sample (*35*) were ligated to the cDNA by mixing 2 μl of 10 μM adapter with 5 μl of cDNA and 1 μl of DMSO and incubating at 75°C for 2 min before placing on ice. Ligation was performed by addition of 15 U T4 RNA ligase in 1× RNA ligase buffer with 18% PEG8000 at 20°C overnight at 1100 rpm. Adapter-ligated cDNA was purified again with MyONE beads before PCR amplification. Initial amplification was performed using 2× Phusion HF Master mix (New England Biolabs, M0531L) with P5Solexa_s and P3Solexa_s for six cycles, followed by ProNex (Promega, NG2001) size selection. Optimal quantitative PCR cycles were determined on a Quantstudio 3 (Thermo Fisher Scientific) using EvaGreen Plus (Biotium, 31077-T), 2× Phusion HF Mastermix, and P5/P3 Solexa primer. Final PCR products were purified using two rounds of ProNex Size selection. Libraries were quantified using a Qubit 4 Fluorometer (Thermo Fisher Scientific) and a TapeStation (Agilent Technologies). Each group of samples was pooled equimolarly and then mixed at the following proportions: 51.7% OAS1 library pool, 38.7% size-matched input library pool, and 8.6% negative control IgG. The pooled pool was sequenced on a NextSeq 550 sequencer with a high-output cartridge. The raw data are available from the Gene Expression Omnibus (GEO) under accession number GSE182394.

iCLIP2 data processing

To identify OAS1-binding sites on human and SARS-CoV-2 RNAs, reads in the raw fastq files from sequencing were demultiplexed to separate samples according to the sample barcodes using Je suite. Sample barcodes and sequencing adapter sequences were trimmed off using cutadapt. Reads were then mapped to a concatenated human (GRCh38, ENSEMBL Release 104) and SARS-CoV-2 (NC_045512.2) genome using STAR with end-to-end alignment mode. The PCR duplicated reads were collapsed to individual reads on the basis of unique molecular identifier barcodes using the Je suite. The GRCh38 and NC_045512.2 annotations were preprocessed with htseq-clip suite to generate sliding windows (Human 50 nt window, 20 nt step size; SARS-CoV-2: 10 nt window, 2 nt step size) (available at https://htseq-clip.readthedocs.io/en/latest/overview.html). Uniquely aligned reads were then used to extract the cross-link truncation site (position 1 relative to the 5′-end of the read start) using bedtools and htseq-clip/extract and quantified using htseq-clip/count. For peak calling, we used the R/Bioconductor DEWSeq package to identify significantly enriched sliding windows in OAS1 immunoprecipitated samples over the corresponding size-matched input control samples (adjusted P value < 0.01, log2-fold change > 2). The IHW R package was used for multiple hypothesis correction. To remove background signal resulting from nonspecific binding of RNA in our immunoprecipitation experiments, sliding windows harboring more cross-link counts (reads per million, log2-fold change > 2) in IgG samples compared with OAS1 immunoprecipitation samples were removed from further analysis. Overlapping significant windows were merged to binding regions, and these sites were curated to 6-nt-long binding peaks on the basis of peak width and maxima using PureCLIP postprocessing R scripts. Binding regions were overlapped with GRCh38 gene annotation (ENSEMBL Release 104) using the GenomicRanges R package. Small noncoding RNA assignment was given priority unless the binding site spanned annotated RNA junctions.

To assess the main driver of differences in the iCLIP samples, we performed a principal components analysis. First, we performed library size correction and variance stabilization transformation implemented in DESeq2. We then used the 1000 sliding windows showing the highest variance to perform principal components analysis using the prcomp function implemented in the base R. For quantification of base composition around the start of the 5′-UTR sequence, reads mapping to the 5′-UTR were extracted using BBMap (https://jgi.doe.gov/data-and-tools/bb-tools/) with kmer = 25 mode. Then the 5′-most base of unique reads was extracted from aligned reads and the composition of bases at the 5′ start site and the 5′ off-template positions were quantified using bam-readcount (https://github.com/genome/bam-readcount). To assess enrichment of iCLIP reads mapping to each chromosome contigs, Samtools/idxstat was used to report alignment summary statistics of deduplicated reads. The ratio of the observed-to-expected read count was calculated using the samtools module of MultiQC software.

To identify enriched motifs within iCLIP hits, 25 nucleotides either side of the peak in signal for each hit were taken as the input. Gene- and region-matched sequences were used as background for ungapped motif prediction. Ungapped motif prediction was performed using both MEME (*40*) and HOMER (*41*) software, and the top predicted motif was selected for each. Gapped motif prediction was performed using GLAM2 (*40*), either allowing for reverse strand search or not. The top predicted motif was selected for each of these.

Synteny analysis

The Ensembl web database was used for assessing the OAS1 locus genome synteny between the human genome (GRCh38.p13) and the available *Rhinolophus* species, *R. ferrumequinum*, genome (mRhiFer1_v1.p). The syntenic region between the human OAS1 exon 7 (ENSE00003913305) and the horseshoe bat genome was examined to identify a region in the latter genome sequence starting at position 7,833,728 of scaffold 25 of the mRhiFer1_v1.p primary assembly that lacked synteny to the human genome. Incidentally, the nonsyntenic region started in-frame where the p46 "CTIL" encoding human sequence would have been. We extracted the 580-bp *R. ferrumequinum* sequence span up to where synteny resumes to the human genome and used hmmscan (HMMER 3.2.1) (*93*) to search against the Dfam database (*94*) for transposable elements present in the sequence. Two confident matches were identified, one to a partial MER74A-like LTR element at the very start of the nonsyntenic sequence and one to a L1-like retrotransposon element at the 3′-end of the sequence. The data and code used for this method are publicly available at the GitHub repository https://github.com/spyros-lytras/bat_OAS1 (*95*).

In silico genome screening

To explore how far back in time this LTR insertion at the OAS1 locus took place, we used Database-Integrated Genome-Screening (DIGS) software (*96*). DIGS uses a nucleotide or amino acid sequence probe to perform a BLAST similarity search through genome assemblies. We collected a set of 44 *Chiroptera* species genome assemblies to perform three in silico screens.

We first used the nucleotide sequence of the syntenic region of *R. ferrumequinum* to human exon 7 (Ensembl) and the adjacent 580-bp region with the detected LTR insertion until homology resumes to the human genome as a probe. The DIGS screen was conducted using a minimum blastn bitscore of 30 and minimum sequence length of 30 nucleotides. Matches were aligned using MAFFT v7.453 and inspected for covering all regions of the probe. The second screen used the CAAX terminal amino acid sequence homologous to that encoded by the human exon 7 of 5 previously annotated bat OAS1 proteins holding a CAAX terminus. The minimum tblastn bitscore was set to 60 and the minimum sequence length to 40 nucleotides. The translated sequences of hits were aligned, and only hits with a CAAX domain present were retained. Finally, to cross-validate that sequence hits of the CAAX domain search are part of the OAS1 locus, we performed a screen using the *R. ferrumequinum* OAS1 C-terminal domain amino acid sequence as a probe, with a minimum tblastn bitscore of 100 and minimum sequence length of 100 nucleotides. Only hits of the CAAX search found on scaffolds with a detected OAS1 locus in the third search were maintained in the analysis. It is worth noting that the lack of an OAS1 domain detected on the same scaffold as hits in the CAAX sequence search is most likely a result of low

genome assembly quality. Regardless, the hits were excluded for completeness. The data and code used for this method are publicly available at the GitHub repository https://github.com/spyros-lytras/bat_OAS1 (*95*).

PDE analysis

To examine the diversity of PDE proteins encoded by coronaviruses, we first constructed an HMMER protein profile. Two seemingly independently acquired PDEs are encoded by the NS2 of Embecoviruses (*62*) and NS4b of MERS-like coronaviruses (*62*), respectively. Group A rotavirus (RVA) has also been described to encode a protein with a homologous PDE domain and similar biological function (*70*). Finally, the AKAP7 mammalian protein holds a PDE domain that has been experimentally shown to complement the function of murine coronaviruses' NS2 activity (*69*). We aligned the amino acid sequence of the PDE domains of the OC43 NS2 (AAT84352.1), the MERS NS4b (AIA22866.1), and the NS4b proteins of two more bat Merbecoviruses HKU5 (YP_001039965.1) and SC2013 (AHY61340.1), the AKAP7 proteins of *Rattus norvegicus* (NP_001001801.1), *Mus musculus* (NP_001366167.1), and humans (NP_057461.2) (as their homology to CoV PDEs has been previously characterized) and the Rotavirus A VP3 protein (AKD32168.1). The alignment was then manually curated using Bioedit on the basis of the homology described in the literature. The final alignment was used to produce an HMM profile using the HMMER suite (v3.2.1) (*93*).

All complete Coronaviridae sequences were downloaded from the NCBI virus online database (*97*). Only sequences with an annotated host and length above 25,000 bp were retained and viruses of "severe acute respiratory syndrome-related coronavirus" species with a human host were excluded, producing a dataset of 2042 complete or near-complete coronavirus genomes. The EMBOSS getorf program was used to extract the translated sequences of all methionine starting ORFs with length >100 nucleotides from the filtered virus genome dataset. All putative ORFs were then screened against our custom PDE HMM profile using hmmscan (*93*). The data and code used for this method are publicly available at the GitHub repository https://github.com/spyros-lytras/bat_OAS1 (*95*).

ISG expression in gastrointestinal and respiratory tissue and an interferome database

The six key cDNAs (UNC93B1, SCARB2, ANKFY1, NCOA7, ZBTB42, and OAS1) that were hits in our screens, SARS-CoV-2 cofactors (ACE2, TMPRSS2, and CSTL), and RNase L were examined for their transcript abundance across respiratory and gastrointestinal tissues using 17,382 RNA-sequencing datasets from the Genotype-Tissue Expression (GTEx) v8 database (GTEx Consortium, 2020). Gene expression is shown as log10 transform of transcripts per million. The respiratory tissue here includes lung and minor salivary gland tissues, and gastrointestinal tissue includes colon, esophagus, and small intestine tissues. To visualize the IFN responsivity in other datasets, an interferome database was used. Data from the Interferome v2.01 (*21*) was downloaded (http://www.interferome.org/). The database was searched for each candidate effector and the following additional search criteria were used: Interferon Type I, Species *Homo Sapiens*, Fold Change Up/Down 1.0. The retrieved experimental data of those genes was downloaded as a text file and used for downstream analysis.

MAIC analysis

The background dataset in MAIC analysis was created as described previously (*10*). MAIC was then run with the human and macaque lists, each independently, to determine the overlap between these lists and the manually curated systematic review of host factors associated with betacoronavirus literature.

Clinical data analysis

A total of 499 whole-blood patient transcriptomes with known disease outcomes were obtained from the ISARIC4C consortium (https://isaric4c.net/). Ethical approval was given by the South Central-Oxford C Research Ethics Committee in England (reference 13/SC/0149) and by the Scotland A Research Ethics Committee (reference 20/SS/0028). The study was registered at https://www.isrctn.com/ISRCTN66726260. Informed consent was obtained from all study participants https://isaric4c.net/protocols/. Underlying data relating to Fig. 5 can be accessed through Edinburgh DataShare (https://doi.org/10.7488/ds/3139);

Preprocessed and STAR (*98*) mapped paired-end reads of 499 whole-blood patient transcriptomes with known disease outcomes from the ISARIC4C study were analyzed to stratify mild (hospitalized but not ICU-admitted patients) and severe (ICU-admitted and/or deceased) patients further into p46-positive and p46-negative groups. Using alignment files as input, strand-specific splice-junction-level counts for each sample were generated using QoRTs (Quality of RNA-seq Tool-Set) (*99*). QoRTs generates a set of nonoverlapping transcript features from the genome annotation, assigns a unique identifier to each feature, and generates counts for each annotated transcript subunit. To validate the method, we applied it to RNA-seq data from IFN-treated A549 cells, which have the AA genotype at Rs10774671 (*29*). We used the samples A549Cas9Clone1, 3, 4, 7, and 10 that were mock treated (NO IFN) or treated with IFN (IFNbeta) using data retrieved from the European Bioinformatics Institute (EBI) under project accession number PRJEB29677. On the basis of the presence or absence of p46 junction counts, mild and severe patient samples were subdivided further into p46-positive (mild and severe) and p46-negative (mild and severe) groups. JunctionSeq (*100*) was used to perform differential usage analysis of both exons and splice junctions using a design model specifying sample and group types. Differential usage results of p46 junction (named J080 by the QoRTs) and the region of exon 5 encoding the C-terminal end of p42 (VRPPASSLPFIPAPLHEA) (named E037 by the QoRTs) were interpreted and visualized using functions of JunctionSeq.

Disease severity and survival were compared in patients with and without expression of the p46 transcript. Severe disease was classified as ICU admission or death, and mild disease as no ICU admission and alive at discharge from hospital. Survival was classified as death or alive at discharge from hospital. Binary logistic regression was used to estimate ORs and 95% CIs with and without adjustment for the effects of age, sex, and ethnicity. Analyses were implemented using IBM SPSS Statistics version 25 (IBM Corp. Armonk, USA). To detect whether one patient group stochastically expressed more of an OAS1 isoform than the other group, we used the nonparametric Mann–Whitney *U* test comparing patients classified as experiencing mild or severe COVID-19. Similarly, where multiple groups were compared (Fig. 4E, right) we used the nonparametric Kruskal–Wallis rank sum test followed by post hoc analysis using the Dunn test. The G frequencies in different populations (at Rs10774671) were extracted from the 1000 genomes project using ensembl. The populations considered were AFR (Africa), AMR (American), EAS (East Asian), EUR (European), and SAS (South Asian). These populations were further subdivided into ASW (African ancestry in SW USA), ACB (African Caribbean in Barbados), BEB (Bengali in Bangladesh), GBR (British from England and Scotland), CDX (Chinese Dai in Xishuangbanna, China), CLM (Colombian in Medellín, Colombia), ESN (Esan in Nigeria), FIN (Finnish in Finland), GWD (Gambian in Western Division – Mandinka), GIH (Gujarati Indians in Houston, Texas, United States), CHB (Han Chinese in Beijing, China), CHS (Han Chinese South, China), IBS (Iberian populations in Spain), ITU (Indian Telugu in the UK) JPT (Japanese in Tokyo, Japan), KHV (Kinh in Ho Chi Minh City, Vietnam), LWK (Luhya in Webuye, Kenya), MSL (Mende in Sierra Leone), MXL (Mexican Ancestry in Los Angeles CA United States), PEL (Peruvian in Lima, Peru), PUR (Puerto Rican in Puerto Rico), PJL (Punjabi in Lahore, Pakistan), STU (Sri Lankan Tamil in the UK), TSI (Toscani in Italia), YRI (Yoruba in Ibadan, Nigeria), and CEU (Utah residents with Northern and

Western European ancestry from the CEPH collection).

REFERENCES AND NOTES

1. E. Pairo-Castineira *et al*., Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021). doi: 10.1038/s41586-020-03065-y; pmid: 33307546
2. P. Bastard *et al*., Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**, eabd4585 (2020). doi: 10.1126/science.abd4585; pmid: 32972996
3. P. D. Monk *et al*., Safety and efficacy of inhaled nebulised interferon beta-1a (SNG001) for treatment of SARS-CoV-2 infection: A randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Respir. Med.* **9**, 196–206 (2021). doi: 10.1016/S2213-2600(20)30511-7; pmid: 33189161
4. R. Channappanavar *et al*., Dysregulated type I interferon and inflammatory monocyte-macrophage responses cause lethal pneumonia in SARS-CoV-infected mice. *Cell Host Microbe* **19**, 181–193 (2016). doi: 10.1016/j.chom.2016.01.007; pmid: 26867177
5. M. Kane *et al*., Identification of interferon-stimulated genes with antiretroviral activity. *Cell Host Microbe* **20**, 392–405 (2016). doi: 10.1016/j.chom.2016.08.005; pmid: 27631702
6. J. W. Schoggins *et al*., A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011). doi: 10.1038/nature09907; pmid: 21478870
7. M. Hoffmann *et al*., SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8 (2020). doi: 10.1016/j.cell.2020.02.052; pmid: 32142651
8. S. J. Rihn *et al*., A plasmid DNA-launched SARS-CoV-2 reverse genetics system and coronavirus toolkit for COVID-19 research. *PLOS Biol.* **19**, e3001091 (2021). doi: 10.1371/journal.pbio.3001091; pmid: 33630831
9. B. G. Hale *et al*., CDK/ERK-mediated phosphorylation of the human influenza A virus NS1 protein at threonine-215. *Virology* **383**, 6–11 (2009). doi: 10.1016/j.virol.2008.10.002; pmid: 19007960
10. N. Parkinson *et al*., Dynamic data-driven meta-analysis for prioritisation of host genes implicated in COVID-19. *Sci. Rep.* **10**, 22303 (2020). doi: 10.1038/s41598-020-79033-3; pmid: 33339864
11. T. Thi Nhu Thao *et al*., Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020). doi: 10.1038/s41586-020-2294-9; pmid: 32365353
12. T. Doyle *et al*., The interferon-inducible isoform of NCOA7 inhibits endosome-mediated viral entry. *Nat. Microbiol.* **3**, 1369–1376 (2018). doi: 10.1038/s41564-018-0273-9; pmid: 30478388
13. J. Donovan, M. Dufner, A. Korennykh, Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1652–1657 (2013). doi: 10.1073/pnas.1218528110; pmid: 23319625
14. Y. Han *et al*., Structure of human RNase L reveals the basis for regulated RNA decay in the IFN response. *Science* **343**, 1244–1248 (2014). doi: 10.1126/science.1249845; pmid: 24578532
15. O. Majer, B. Liu, L. S. M. Kreuk, N. Krogan, G. M. Barton, UNC93B1 recruits syntenin-1 to dampen TLR7 signalling and prevent autoimmunity. *Nature* **575**, 366–370 (2019). doi: 10.1038/s41586-019-1612-6; pmid: 31546246
16. S. Yamayoshi *et al*., Scavenger receptor B2 is a cellular receptor for enterovirus 71. *Nat. Med.* **15**, 798–801 (2009). doi: 10.1038/nm.1992; pmid: 19543282
17. S. Heybrock *et al*., Lysosomal integral membrane protein-2 (LIMP-2/SCARB2) is involved in lysosomal cholesterol export. *Nat. Commun.* **10**, 3521 (2019). doi: 10.1038/s41467-019-11425-0; pmid: 31387993
18. D. Neculai *et al*., Structure of LIMP-2 provides functional insights with implications for SR-BI and CD36. *Nature* **504**, 172–176 (2013). doi: 10.1038/nature12684; pmid: 24162852
19. T. P. Peacock *et al*., The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat. Microbiol.* **6**, 899–909 (2021). doi: 10.1038/s41564-021-00908-w; pmid: 33907312
20. H. Khan, H. Winstone, J. Jimenez-Guardeno, C. Graham, K. J. Doores, C. Goujon, D. A. Matthews, A. D. Davidson, S. J. Rihn, M. Palmarini, S. J. D. Neil, M. H. Malim, TMPRSS2 promotes SARS-CoV-2 evasion from NCOA7-mediated restriction. bioRxiv 453488 [Preprint] (2021). doi: 10.1101/2021.07.23.453488
21. I. Rusinova *et al*., Interferome v2.0: An updated database of annotated interferon-regulated genes. *Nucleic Acids Res.* **41**, D1040–D1046 (2013). doi: 10.1093/nar/gks1215; pmid: 23203888
22. S. Zhou *et al*., A Neanderthal OAS1 isoform protects individuals of European ancestry against COVID-19 susceptibility and severity. *Nat. Med.* **27**, 659–667 (2021). doi: 10.1038/s41591-021-01281-1; pmid: 33633408
23. Y. J. Hou *et al*., SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell* **182**, 429–446.e14 (2020). doi: 10.1016/j.cell.2020.05.042; pmid: 32526206
24. H. Zeberg, S. Pääbo, A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026309118 (2021). doi: 10.1073/pnas.2026309118; pmid: 33593941
25. COVID-19 Host Genetics Initiative, Mapping the human genetic architecture of COVID-19. *Nature* (2021). pmid: 34237774
26. A. E. Shaw *et al*., Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLOS Biol.* **15**, e2004086 (2017). doi: 10.1371/journal.pbio.2004086; pmid: 29253856
27. A. G. Hovanessian, R. E. Brown, I. M. Kerr, Synthesis of low molecular weight inhibitor of protein synthesis with enzyme from interferon-treated cells. *Nature* **268**, 537–540 (1977). doi: 10.1038/268537a0; pmid: 560630
28. C. Bisbal, R. H. Silverman, Diverse functions of RNase L and implications in pathology. *Biochimie* **89**, 789–798 (2007). doi: 10.1016/j.biochi.2007.02.006; pmid: 17400356
29. Y. Li *et al*., Activation of RNase L is dependent on OAS3 expression during infection with diverse human viruses. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2241–2246 (2016). doi: 10.1073/pnas.1519657113; pmid: 26858407
30. J. M. Burke, S. L. Moon, T. Matheny, R. Parker, RNase L reprograms translation by widespread mRNA turnover escaped by antiviral mRNAs. *Mol. Cell* **75**, 1203–1217.e5 (2019). doi: 10.1016/j.molcel.2019.07.029; pmid: 31494035
31. S. Rath *et al*., Concerted 2-5A-mediated mRNA decay and transcription reprogram protein synthesis in the dsRNA response. *Mol. Cell* **75**, 1218–1228.e6 (2019). doi: 10.1016/j.molcel.2019.07.027; pmid: 31494033
32. A. Zhou *et al*., Interferon action and apoptosis are defective in mice devoid of 2′,5′-oligoadenylate-dependent RNase L. *EMBO J.* **16**, 6355–6363 (1997). doi: 10.1093/emboj/16.21.6355; pmid: 9351818
33. K. Malathi, B. Dong, M. Gale Jr., R. H. Silverman, Small self-RNA generated by RNase L amplifies antiviral innate immunity. *Nature* **448**, 816–819 (2007). doi: 10.1038/nature06042; pmid: 17653195
34. A. Quintás-Cardama *et al*., Preclinical characterization of the selective JAK1/2 inhibitor INCB018424: Therapeutic implications for the treatment of myeloproliferative neoplasms. *Blood* **115**, 3109–3117 (2010). doi: 10.1182/blood-2009-04-214957; pmid: 20130243
35. A. Buchbender *et al*., Improved library preparation with the new iCLIP2 protocol. *Methods* **178**, 33–48 (2020). doi: 10.1016/j.ymeth.2019.10.003; pmid: 31610236
36. E. L. Van Nostrand *et al*., Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016). doi: 10.1038/nmeth.3810; pmid: 27018577
37. Z. Miao, A. Tidu, G. Eriani, F. Martin, Secondary structure of the SARS-CoV-2 5′-UTR. *RNA Biol.* **18**, 447–456 (2021). doi: 10.1080/15476286.2020.1814556; pmid: 32965173
38. M. Garcia-Moreno *et al*., System-wide profiling of RNA-binding proteins uncovers key regulators of virus infection. *Mol. Cell* **74**, 196–211.e11 (2019). doi: 10.1016/j.molcel.2019.01.017; pmid: 30799147
39. A. Dhir *et al*., Mitochondrial double-stranded RNA triggers antiviral signalling in humans. *Nature* **560**, 238–242 (2018). doi: 10.1038/s41586-018-0363-0; pmid: 30046113
40. T. L. Bailey *et al*., MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37** (Web Server), W202–W208 (2009). doi: 10.1093/nar/gkp335; pmid: 19458158
41. S. Heinz *et al*., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010). doi: 10.1016/j.molcel.2010.05.004; pmid: 20513432
42. S. L. Schwartz *et al*., Human OAS1 activation is highly dependent on both RNA sequence and context of activating RNA motifs. *Nucleic Acids Res.* **48**, 7520–7531 (2020). pmid: 32678884
43. R. Kodym, E. Kodym, M. D. Story, 2′-5′-Oligoadenylate synthetase is activated by a specific RNA sequence motif. *Biochem. Biophys. Res. Commun.* **388**, 317–322 (2009). doi: 10.1016/j.bbrc.2009.07.167; pmid: 19665006
44. C. E. Melia *et al*., The origin, dynamic morphology, and PI4P-independent formation of encephalomyocarditis virus replication organelles. *mBio* **9**, e00420-18 (2018). doi: 10.1128/mBio.00420-18; pmid: 29666283
45. J. Chebath, P. Benech, M. Revel, M. Vigneron, Constitutive expression of (2′-5′) oligo A synthetase confers resistance to picornavirus infection. *Nature* **330**, 587–588 (1987). doi: 10.1038/330587a0; pmid: 2825034
46. M. Drappier, T. Michiels, Inhibition of the OAS/RNase L pathway by viruses. *Curr. Opin. Virol.* **15**, 19–26 (2015). doi: 10.1016/j.coviro.2015.07.002; pmid: 26231767
47. J. Y. Min, R. M. Krug, The primary function of RNA binding by the influenza A virus NS1 protein in infected cells: Inhibiting the 2′-5′ oligo (A) synthetase/RNase L pathway. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7100–7105 (2006). doi: 10.1073/pnas.0602184103; pmid: 16627618
48. K. Knoops *et al*., SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLOS Biol.* **6**, e226 (2008). doi: 10.1371/journal.pbio.0060226; pmid: 18798692
49. I. Romero-Brey, R. Bartenschlager, Endoplasmic reticulum: The favorite intracellular niche for viral replication and assembly. *Viruses* **8**, 160 (2016). doi: 10.3390/v8060160; pmid: 27338443
50. E. J. Snijder *et al*., Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. *J. Virol.* **80**, 5927–5940 (2006). doi: 10.1128/JVI.02501-05; pmid: 16731931
51. V. Bonnevie-Nielsen *et al*., Variation in antiviral 2′,5′-oligoadenylate synthetase (2′5′AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene. *Am. J. Hum. Genet.* **76**, 623–633 (2005). doi: 10.1086/429391; pmid: 15732009
52. H. Li *et al*., Identification of a Sjögren's syndrome susceptibility locus at OAS1 that influences isoform switching, protein expression, and responsiveness to type I interferons. *PLOS Genet.* **13**, e1006820 (2017). doi: 10.1371/journal.pgen.1006820; pmid: 28640813
53. S. Noguchi *et al*., Differential effects of a common splice site polymorphism on the generation of OAS1 variants in human bronchial epithelial cells. *Hum. Immunol.* **74**, 395–401 (2013). doi: 10.1016/j.humimm.2012.11.011; pmid: 23220500
54. J. K. Lim *et al*., Genetic variation in OAS1 is a risk factor for initial infection with West Nile virus in man. *PLOS Pathog.* **5**, e1000321 (2009). doi: 10.1371/journal.ppat.1000321; pmid: 19247438
55. M. K. El Awady *et al*., Single nucleotide polymorphism at exon 7 splice acceptor site of OAS1 gene determines response of hepatitis C virus patients to interferon therapy. *J. Gastroenterol. Hepatol.* **26**, 843–850 (2011). doi: 10.1111/j.1440-1746.2010.06605.x; pmid: 21182542
56. C. M. Carey *et al*., Recurrent loss-of-function mutations reveal costs to OAS1 antiviral activity in primates. *Cell Host Microbe* **25**, 336–343.e4 (2019). doi: 10.1016/j.chom.2019.01.001; pmid: 30713099
57. S. Maurer-Stroh, F. Eisenhaber, Refinement and prediction of protein prenylation motifs. *Genome Biol.* **6**, R55 (2005). doi: 10.1186/gb-2005-6-6-r55; pmid: 15960807
58. S. Skrivergaard *et al*., The cellular localization of the p42 and p46 oligoadenylate synthetase 1 isoforms and their impact on mitochondrial respiration. *Viruses* **11**, 1122 (2019). doi: 10.3390/v11121122; pmid: 31817188
59. J. Huffman, G. Butler-Laporte, A. Khan, T. G. Drivas, G. M. Peloso, T. Nakanishi, A. Verma, K. Kiryluk, J. B. Richards, H. Zeberg, Alternative splicing of OAS1 alters the risk for severe COVID-19. medRxiv 2021.03.20.21254005 [Preprint] (2021). doi: 10.1101/2021.03.20.21254005
60. L. Zhao *et al*., Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. *Cell Host Microbe* **11**, 607–616 (2012). doi: 10.1016/j.chom.2012.04.011; pmid: 22704621
61. S. A. Goldstein *et al*., Lineage A betacoronavirus NS2 proteins and the homologous torovirus berne pp1a carboxy-terminal domain are phosphodiesterases that antagonize activation of RNase L. *J. Virol.* **91**, e02201-16 (2017). doi: 10.1128/JVI.02201-16; pmid: 28003490
62. J. M. Thornbrough *et al*., Middle East respiratory syndrome coronavirus NS4b protein inhibits host RNase L activation.

mBio **7**, e00258 (2016). doi: 10.1128/mBio.00258-16; pmid: 27025250
63. S. K. Lau *et al*., Discovery of a novel coronavirus, China Rattus coronavirus HKU24, from Norway rats supports the murine origin of Betacoronavirus 1 and has implications for the ancestor of Betacoronavirus lineage A. *J. Virol.* **89**, 3076–3092 (2015). doi: 10.1128/JVI.02420-14; pmid: 25552712
64. L. Vijgen *et al*., Complete genomic sequence of human coronavirus OC43: Molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J. Virol.* **79**, 1595–1604 (2005). doi: 10.1128/JVI.79.3.1595-1604.2005; pmid: 15650185
65. A. Moreno *et al*., Detection and full genome characterization of two beta CoV viruses related to Middle East respiratory syndrome from bats in Italy. *Virol. J.* **14**, 239 (2017). doi: 10.1186/s12985-017-0907-1; pmid: 29258555
66. C. B. Reusken *et al*., Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: A comparative serological study. *Lancet Infect. Dis.* **13**, 859–866 (2013). doi: 10.1016/S1473-3099(13)70164-6; pmid: 23933067
67. T. Briese *et al*., Middle East respiratory syndrome coronavirus quasispecies that include homologues of human isolates revealed through whole-genome analysis and virus cultured from dromedary camels in Saudi Arabia. *mBio* **5**, e01146–e14 (2014). doi: 10.1128/mBio.01146-14; pmid: 24781747
68. S. Lytras, J. Hughes, D. Martin, A. de Klerk, R. Lourens, S. L. Kosakovsky Pond, W. Xia, X. Jiang, D. L. Robertson, Exploring the natural origins of SARS-CoV-2 in the light of recombination. bioRxiv 427830 [Preprint] (2021). doi: 10.1101/2021.01.22.427830
69. E. Gusho *et al*., Murine AKAP7 has a 2′,5′-phosphodiesterase domain that can complement an inactive murine coronavirus ns2 gene. *mBio* **5**, e01312–e01314 (2014). doi: 10.1128/mBio.01312-14; pmid: 24987090
70. R. Zhang *et al*., Homologous 2′,5′-phosphodiesterases from disparate RNA viruses antagonize antiviral innate immunity. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13114–13119 (2013). doi: 10.1073/pnas.1306917110; pmid: 23878220
71. X. Y. Ge *et al*., Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013). doi: 10.1038/nature12711; pmid: 24172901
72. Y. Guan *et al*., Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276–278 (2003). doi: 10.1126/science.1087139; pmid: 12958366
73. O. A. MacLean *et al*., Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. *PLOS Biol.* **19**, e3001115 (2021). doi: 10.1371/journal.pbio.3001115; pmid: 33711012
74. E. K. Tsinda, G. S. Mmbando, Recent updates on the possible reasons for the low incidence and morbidity of COVID-19 cases in Africa. *Bull. Natl. Res. Cent.* **45**, 133 (2021). doi: 10.1186/s42269-021-00589-9; pmid: 34335014
75. Centers for Disease Control and Prevention, "COVID-19: Hospitalization and death by race/ethnicity" (CDC, 2021); https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html.
76. L. L. Field *et al*., OAS1 splice site polymorphism controlling antiviral enzyme activity influences susceptibility to type 1 diabetes. *Diabetes* **54**, 1588–1591 (2005). doi: 10.2337/diabetes.54.5.1588; pmid: 15855350
77. K. J. Olival *et al*., Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017). doi: 10.1038/nature22975; pmid: 28636590
78. A. Banerjee *et al*., Novel insights into immune systems of bats. *Front. Immunol.* **11**, 26 (2020). doi: 10.3389/fimmu.2020.00026; pmid: 32117225
79. N. B. Simmons, A. L. Cirranello, "Bat species of the world: A taxonomic and geographic database" (2020); https://batnames.org/.
80. N. Mollentze, D. G. Streicker, Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9423–9430 (2020). doi: 10.1073/pnas.1919176117; pmid: 32284401
81. J. Xie *et al*., Dampened STING-dependent interferon activation in bats. *Cell Host Microbe* **23**, 297–301.e4 (2018). doi: 10.1016/j.chom.2018.01.006; pmid: 29478775
82. Y. C. Kwon, J. I. Kang, S. B. Hwang, B. Y. Ahn, The ribonuclease L-dependent antiviral roles of human 2′,5′-oligoadenylate synthetase family members against hepatitis C virus. *FEBS Lett.* **587**, 156–164 (2013). doi: 10.1016/j.febslet.2012.11.010; pmid: 23196181
83. J. W. Schoggins *et al*., Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature* **505**, 691–695 (2014). doi: 10.1038/nature12862; pmid: 24284630
84. J. Zhao *et al*., 2′,5′-Oligoadenylate synthetase 1(OAS1) inhibits PRRSV replication in Marc-145 cells. *Antiviral Res.* **132**, 268–273 (2016). doi: 10.1016/j.antiviral.2016.07.001; pmid: 27395032
85. W. Wang *et al*., Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. *Virology* **474**, 19–27 (2015). doi: 10.1016/j.virol.2014.10.017; pmid: 25463600
86. S. J. Rihn *et al*., TRIM69 inhibits vesicular stomatitis Indiana virus. *J. Virol.* **93**, e00951-19 (2019). doi: 10.1128/JVI.00951-19; pmid: 31375575
87. E. de Wit *et al*., Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments. *Virus Res.* **103**, 155–161 (2004). doi: 10.1016/j.virusres.2004.02.028; pmid: 15163504
88. M. L. Turnbull *et al*., Role of the B allele of influenza A virus segment 8 in setting mammalian host range and pathogenicity. *J. Virol.* **90**, 9263–9284 (2016). doi: 10.1128/JVI.01205-16; pmid: 27489273
89. L. K. Hallak, P. L. Collins, W. Knudson, M. E. Peeples, Iduronic acid-containing glycosaminoglycans on target cells are required for efficient respiratory syncytial virus infection. *Virology* **271**, 264–275 (2000). doi: 10.1006/viro.2000.0293; pmid: 10860881
90. J. Feng *et al*., Interferon-stimulated gene (ISG)-expression screening reveals the specific antibunyaviral activity of ISG20. *J. Virol.* **92**, e02140-17 (2018). doi: 10.1128/JVI.02140-17; pmid: 29695422
91. C. E. Stewart, R. E. Randall, C. S. Adamson, Inhibitors of the interferon response enhance virus replication in vitro. *PLOS ONE* **9**, e112014 (2014). doi: 10.1371/journal.pone.0112014; pmid: 25390891
92. B. J. Zarnegar *et al*., irCLIP platform for efficient characterization of protein-RNA interactions. *Nat. Methods* **13**, 489–492 (2016). doi: 10.1038/nmeth.3840; pmid: 27111506
93. S. R. Eddy, A new generation of homology search tools based on probabilistic inference. *Genome Inform.* **23**, 205–211 (2009). doi: 10.1142/9781848165632_0019; pmid: 20180275
94. R. Hubley *et al*., The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44** (D1), D81–D89 (2016). doi: 10.1093/nar/gkv1272; pmid: 26612867
95. S. Lytras, "OAS1 bat analysis" (2021); doi: 10.5281/zenodo.5513932
96. H. Zhu, T. Dennis, J. Hughes, R. J. Gifford, Database-integrated genome screening (DIGS): exploring genomes heuristically using sequence similarity search tools and a relational database. *bioRxiv*, 246835 (2018). doi: 10.1101/246835
97. E. L. Hatcher *et al*., Virus Variation Resource - improved response to emergent viral outbreaks. *Nucleic Acids Res.* **45** (D1), D482–D490 (2017). doi: 10.1093/nar/gkw1065; pmid: 27899678
98. A. Dobin *et al*., STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi: 10.1093/bioinformatics/bts635; pmid: 23104886
99. S. W. Hartley, J. C. Mullikin, QoRTs: A comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics* **16**, 224 (2015). doi: 10.1186/s12859-015-0670-5; pmid: 26187896
100. S. W. Hartley, J. C. Mullikin, Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. *Nucleic Acids Res.* **44**, e127 (2016). doi: 10.1093/nar/gkw501; pmid: 27257077
101. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017). doi: 10.1093/molbev/msx116; pmid: 28387841

ACKNOWLEDGMENTS

This report is independent research that used data provided by the Medical Research Council–funded ISARIC4C Consortium, which the Consortium collected under a research contract funded by the National Institute for Health Research (NIHR). The views expressed in this publication are those of the author(s) and not necessarily those of the ISARIC4C consortium. We thank P. Moss and the UK Coronavirus Immunology Consortium (UK-CIC) for their guidance and suggestions; V. Thiel, P. Collins, M. Peiris, B. Haagmans, A. Patel, A. Kohl, S. Neil, M. Bouloy, W. Barclay, and the G2P-UK consortium for viruses and cells; A. Aziz, R. Gifford, A. Szemiel, and W. Furnon for technical assistance; Nexcelom (Celigo) for their technical assistance and support during the UK COVID-19 lockdowns (in 2020); J. J. Farrar and N. Shindo for support. This work used data provided by patients and collected by the National Health Service (NHS) as part of their care and support #DataSavesLives. We are grateful to the 2648 frontline NHS clinical and research staff and volunteer medical students who collected the data in challenging circumstances and to the participants and their families for their generous individual contributions in these difficult times. SARS-CoV-2 England/2/2020 was obtained under a materials transfer agreement (MTA) with Public Health England, SynSARS-CoV-2-eGFP was obtained under an MTA with the University of Bern, Rift Valley fever phlebovirus replicon cells and glycoprotein expression plasmid were obtained under an MTA with Stichting Wageningen Research, and Chandipura vesiculovirus EGFP was obtained under an MTA from the University of Warwick. SARS-CoV-2 CVR-GLA-1, SARS-CoV-2-ZsGreen in addition to the modified cell lines (described herein) and plasmids used for these modifications are available from S.J.W. under an MTA with the University of Glasgow. **Funding:** This work was partly funded by UKRI/NIHR through the UK Coronavirus Immunology Consortium (UK-CIC MR/V028448/1 to M.P. and S.J.W.) and the MRC through the following grants: MR/ K024752/1 (to S.J.W.), MC_UU_12014/10 (to M.P. and S.J.W.), MC_UU_12014/12 (to J.H., D.L.R., and S.K.) and MR/P022642/1 (to S.J.W. and S.J.R.), MR/V000489/1 (to E.C.Y.W. and R.J.S.), MR/S00971X/1 (to R.J.S. and E.C.Y.W.), MR/P001602/1 (to E.C.Y.W.), and MR/V011561/1 (to P.J.L.). Support was also provided by a Wellcome Principal Research Fellowship 210688/Z/18/Z (to P.J.L.), a Wellcome Trust Fellowship 201366/Z/16/Z (to S.J.R.), a Wellcome Investigator Award 209412/Z/17/Z (to I.D.), the Addenbrooke's Charitable Trust and the NIHR Cambridge Biomedical Research Centre (to P.J.L.), support from the German Research Foundation, Deutsche Forschungsgemeinschaft; project number 406109949 (to V.H.), and German Federal Ministry of Food and Agriculture through BMEL Förderkennzeichen: 01KI1723G (to V.H.), and a Daphne Jackson Fellowship funded by Medical Research Scotland (to S.S.). A.C. is supported by MRC grants MR/R021562/1, MC_UU_12014/10, and MC_UU_12014/12. J.Y.L. is funded by a Medial Sciences Graduate Studentship, University of Oxford. L.I. is funded by BBSRC DTP scholarship number BB/M011224/1. MAIC analysis was supported by the SHIELD Consortium (MRC grant MRNO2995X/1). ISARIC4C is supported by grants from the Medical Research Council (grant MC_PC_19059), the NIHR (award CO-CIN-01) and by the NIHR Health Protection Research Unit (HPRU) in Emerging and Zoonotic Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with Liverpool School of Tropical Medicine and the University of Oxford (award 200907), NIHR HPRU in Respiratory Infections at Imperial College London with PHE (award 200927), Wellcome Trust and Department for International Development (215091/Z/18/Z), and the Bill and Melinda Gates Foundation (OPP1209135), and Liverpool Experimental Cancer Medicine Centre (C18616/A25153), NIHR Biomedical Research Centre at Imperial College London (IS-BRC-1215-20013), EU Platform foR European Preparedness Against (Re-)emerging Epidemics (PREPARE) (FP7 project 602525) and NIHR Clinical Research Network provided infrastructure support for this research. **Author contributions:** A.W. and E.S. led the molecular virology work. S.L. guided the genomic data analysis and S.K. directed the clinical data analysis. M.L.T. headed the immunofluorescence work. M.N., A.W., and A.C. designed the iCLIP2 experiments. M.N., N.P., and A.W. performed the iCLIP2 experiment. J.Y.L. and L.I. analysed the iCLIP2 data. J.Y.L., L.I., M.N., I.D., A.C., and S.J.W. interpreted the iCLIP2 results. Conceptualization: J.K.B., M.G.S. and P.J.M.O. led and created the ISARIC4C consortium and A.W., A.C., M.P., P.J.L., S.J.R., and S.J.W. conceived of this study. Data curation: S.L., S.K., M.N., C.L., V.H., J.Y.L. Formal Analysis: A.W., E.S., S.L., S.K., M.N., M.L.T., C.L., J.Y.L., L.I., S.C., B.W., R.F.J., A.C., S.J.W. Funding acquisition: R.J.S., E.C.Y.W., I.D., A.C., D.L.R., M.G.S., P.J.M.O., M.P., P.J.L., J.K.B., S.J.R., S.J.W. Investigation: A.W., E.S., S.L., M.N., M.L.T., C.L., V.H., J.A., I.J., N.C.R., M.V., R.M.P., J.Y.L., L.I., N.P., D.G.S., S.S., E.J.D.G., T.W.M.C., Q.G., E.L.D., S.C., B.W., F.T.M.C., M.F.S., L.C.L.F., M.P., S.J.R. Methodology: A.W., S.L., M.N., C.L., M.V., S.S. Project administration: S.J.W. Resources: A.W., S.K., E.J.D.G., T.W.M.C., F.T.M.C., M.F.S., L.C.L.F., L.M., A.F., A.M., G.G., J.L.S.F., M.M., A.H., M.G.S., P.J.M.O., J.K.B. Software: S.K., J.Y.L., L.I. Supervision: V.H., J.H., I.D., A.C., D.L.R., M.P., P.J.L., J.K.B., S.J.R., S.J.W. Validation: A.W., M.N., V.H., F.T.M.C. Visualization: A.W., E.S., S.L., S.K., M.L.T., C.L., V.H., J.Y.L., Q.G., S.J.W. Writing – original draft: A.W., E.S., S.L., S.K., M.N., A.C., S.J.W. Writing – review and editing: A.W., E.S., S.L., S.K.,

M.N., M.L.T., C.L., V.H., J.A., I.J., N.C.R., M.V., R.M.P., J.Y.L., L.I., N.P., D.G.S., S.S., E.J.D.G., T.W.M.C., Q.G., E.L.D., S.C., B.W., F.T.M.C., M.F.S., L.C.L.F., J.L.S.F., M.M., J.H., R.J.S., E.C.Y.W., A.H., I.D., R.F.J., A.C., D.L.R., M.G.S., P.J.M.O., M.P., P.J.L., J.K.B., S.J.R., S.J.W.
Competing interests: The authors declare no competing financial interests. **Data and materials availability:** The raw iCLIP sequencing data are available from GEO under accession number GSE182394. The data and code used for genome analysis are publicly available at the GitHub repository https://github.com/spyros-lytras/bat_OAS1 and permanently on Zenodo (*95*). The ISARIC WHO CCP-UK study protocol is available at https://isaric4c.net/protocols; study registry https://www.isrctn.com/ISRCTN66726260. Underlying data relating to Fig. 5 can be accessed through Edinburgh DataShare (https://doi.org/10.7488/ds/3139). All other underlying data are available from Enlighten (https://doi.org/10.5525/gla.researchdata.1178). 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.org/doi/10.1126/science.abj3624
Figs. S1 to S7
Table S1
ISARIC4C Author List
MDAR Reproducibility Checklist

7 May 2021; accepted 23 September 2021
Published online 28 September 2021
10.1126/science.abj3624

RESEARCH ARTICLES

LIPID TRANSPORT

Distinct allosteric mechanisms of first-generation MsbA inhibitors

François A. Thélot[1,2], Wenyi Zhang[3,4], KangKang Song[5,6], Chen Xu[5,6], Jing Huang[3,4], Maofu Liao[1]*

ATP-binding cassette (ABC) transporters couple adenosine 5′-triphosphate (ATP) hydrolysis to substrate transport across biological membranes. Although many are promising drug targets, their mechanisms of modulation by small-molecule inhibitors remain largely unknown. Two first-generation inhibitors of the MsbA transporter, tetrahydrobenzothiophene 1 (TBT1) and G247, induce opposite effects on ATP hydrolysis. Using single-particle cryo–electron microscopy and functional assays, we show that TBT1 and G247 bind adjacent yet separate pockets in the MsbA transmembrane domains. Two TBT1 molecules asymmetrically occupy the substrate-binding site, which leads to a collapsed inward-facing conformation with decreased distance between the nucleotide-binding domains (NBDs). By contrast, two G247 molecules symmetrically increase NBD distance in a wide inward-open state of MsbA. The divergent mechanisms of action of these MsbA inhibitors provide important insights into ABC transporter pharmacology.

ATP-binding cassette (ABC) transporters perform diverse functions that include importing and exporting small molecules, mediating ion channel opening, and translocating lipids across cell membranes (*1*–*8*). ABC transporters have two transmembrane domains (TMDs), which interact with and transport substrates, and two nucleotide-binding domains (NBDs), which bind and hydrolyze adenosine 5′-triphosphate (ATP). Decades of investigation have led to various models to describe ABC transporter mechanisms. In the alternating access model (*9*–*13*), the substrate is first recognized by the inward-facing transporter, subsequent ATP binding in the NBDs promotes a transition to the outward-facing state and substrate release, and last, ATP hydrolysis resets the transporter in the inward-facing state. Interrupting such conformational transition cycles has direct applications in treating cancer, regulating cholesterol homeostasis, and developing antibiotics (*14*–*16*).

Small-molecule inhibitors have been developed against human multidrug ABC transporters, including the ABCB1 inhibitors zosuquidar, tariquidar, and elacridar and the ABCG2 inhibitors MZ29 and MB136 (*3*, *17*–*19*). Structural studies have revealed how these compounds bind to the TMDs, which interrupts conformational transition and consequently decreases adenosine triphosphatase (ATPase) activity (*3*, *19*, *20*). By contrast, there is little structural insight into small-molecule modulation of most other ABC transporters, which demonstrate narrower substrate specificity. We do not know whether modes of small-molecule inhibition are shared across the ABC superfamily or whether there are general druggable conformations or pockets. These knowledge gaps limit our ability to rationally design drugs that target many ABC transporters.

MsbA is a model system to study ABC transporter mechanism (*21*, *22*). Because of its essential role in lipopolysaccharide (LPS) biogenesis in Gram-negative bacteria, MsbA is also an attractive target for developing antibiotics (*23*). X-ray crystallography and cryo–electron microscopy (cryo-EM) studies have determined MsbA structures in several conformations, including ligand free, LPS bound, adenylyl-imidodiphosphate (AMP-PNP) bound, and vanadate trapped (*6*, *22*, *24*–*26*). The structural basis of specific binding and transport of LPS by MsbA has been elucidated (*6*, *25*, *27*). Compared with the recent rapid progress in studying MsbA function, the understanding of MsbA inhibition has lagged behind (*27*).

Two types of specific MsbA inhibitors have been reported: One has a tetrahydrobenzothiophene (TBT) scaffold, named here as TBT1, and

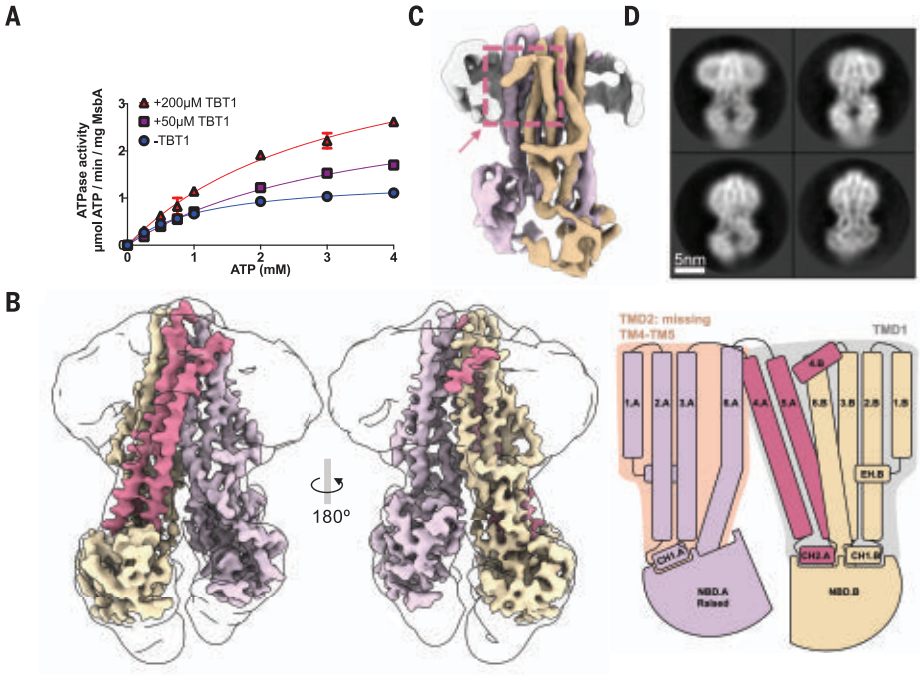


Fig. 1. TBT1 binding induces a collapsed inward-facing conformation of MsbA. (**A**) ATPase activity of *A. baumannii* MsbA in nanodiscs, measured at varying ATP and TBT1 concentrations. TBT1 increases the rate of ATP hydrolysis in a dose-dependent manner. Each point represents mean ± SD (calculated from three independent measurements). (**B**) (Left) Cryo-EM map (4.0-Å resolution) of *A. baumannii* MsbA in complex with TBT1, with domain-swapping TM4–TM5 colored fuchsia. The unsharpened map filtered at 10-Å resolution is displayed as an outline to show the nanodisc and tightened NBD positioning. (Right) Cartoon of TBT1-bound MsbA with structural elements indicated that illustrates the structural asymmetry of the two MsbA chains. EH, elbow helix. A flipped cartoon can be found in fig. S4A. (**C**) Cryo-EM reconstruction low-pass filtered at 6-Å resolution. The dashed box points to the N-terminal end of TM4.B, where TM4–TM5.B becomes disordered and absent from the reconstruction. (**D**) Representative 2D class averages of TBT1-bound MsbA, showing constricted TMDs yet separate NBDs. The box size is 203 Å.

[1]Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA. [2]Biological and Biomedical Sciences Program, Harvard University, Cambridge, MA, USA. [3]Key Laboratory of Structural Biology of Zhejiang Province, Westlake University, Hangzhou, China. [4]Westlake AI Therapeutics Lab, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China. [5]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA. [6]Cryo-EM Core Facility, University of Massachusetts Medical School, Worcester, MA, USA.
*Corresponding author. Email: maofu_liao@hms.harvard.edu

stimulates MsbA ATPase activity despite abolishing LPS transport (*28*); the other group, here referred to as G compounds, blocks both ATP hydrolysis and LPS transport (*25*, *29*). Although these first-generation MsbA inhibitors represent an important breakthrough, their mechanisms of inhibition are not well understood. Particularly, TBT1-induced decoupling of ATPase activity from substrate transport is puzzling because most known inhibitors suppress ATP hydrolysis of ABC transporters in membrane environment (*12*). We thus sought to understand how TBT1 and G compounds exert opposite allosteric effects on ATP hydrolysis, even though both block LPS transport.

TBT1-bound MsbA adopts an asymmetric, collapsed inward-facing conformation

TBT1 was identified as an LPS transport inhibitor and MsbA ATPase stimulator in strains from the *Acinetobacter* genus (*28*). We therefore sought to explore the mechanism of TBT1 inhibition in *Acinetobacter baumannii*, an ESKAPE pathogen (for *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *A. baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species) responsible for antibiotic-resistant infections in patients (*30*). *A. baumannii* MsbA was expressed in *Escherichia coli* cells, purified in dodecyl maltoside (DDM), and reconstituted in palmitoyl-oleoyl-phosphatidylglycerol nanodiscs (fig. S1, A to G) (*31*, *32*). The basal ATPase activity of *A. baumannii* MsbA in nanodiscs was ~1 μmol ATP per minute per milligram of MsbA (Fig. 1A), which is approximately fourfold lower than that of *E. coli* MsbA (*6*). Exposure of MsbA to TBT1 resulted in four- to sixfold stimulation in ATPase activity (figs. S1, D and G, and S7B), as well as a dose-dependent increase in reaction rate in Michaelis-Menten kinetics (Fig. 1A). TBT1-induced ATPase stimulation was much more pronounced in nanodiscs than in DDM (fig. S1G), which suggests that nanodiscs are a more native-like system than detergent for investigating TBT1 action on MsbA. We thus used single-particle cryo-EM to determine the structure of TBT1-bound MsbA in nanodiscs at an overall resolution of 4.3 Å, with the TMDs at 4.0-Å resolution showing sufficient side-chain densities for model building (Fig. 1B and figs. S2 and S3).

The conformation of TBT1-bound MsbA is asymmetric and distinct from all previously determined structures of type IV ABC transporters, which are characterized by their domain-swapped transmembrane helices (TMs) (*33*). The domain-swapping TM4–TM5 bundle is well resolved in chain A yet disordered in chain B (Fig. 1B), which is also evident in unmasked

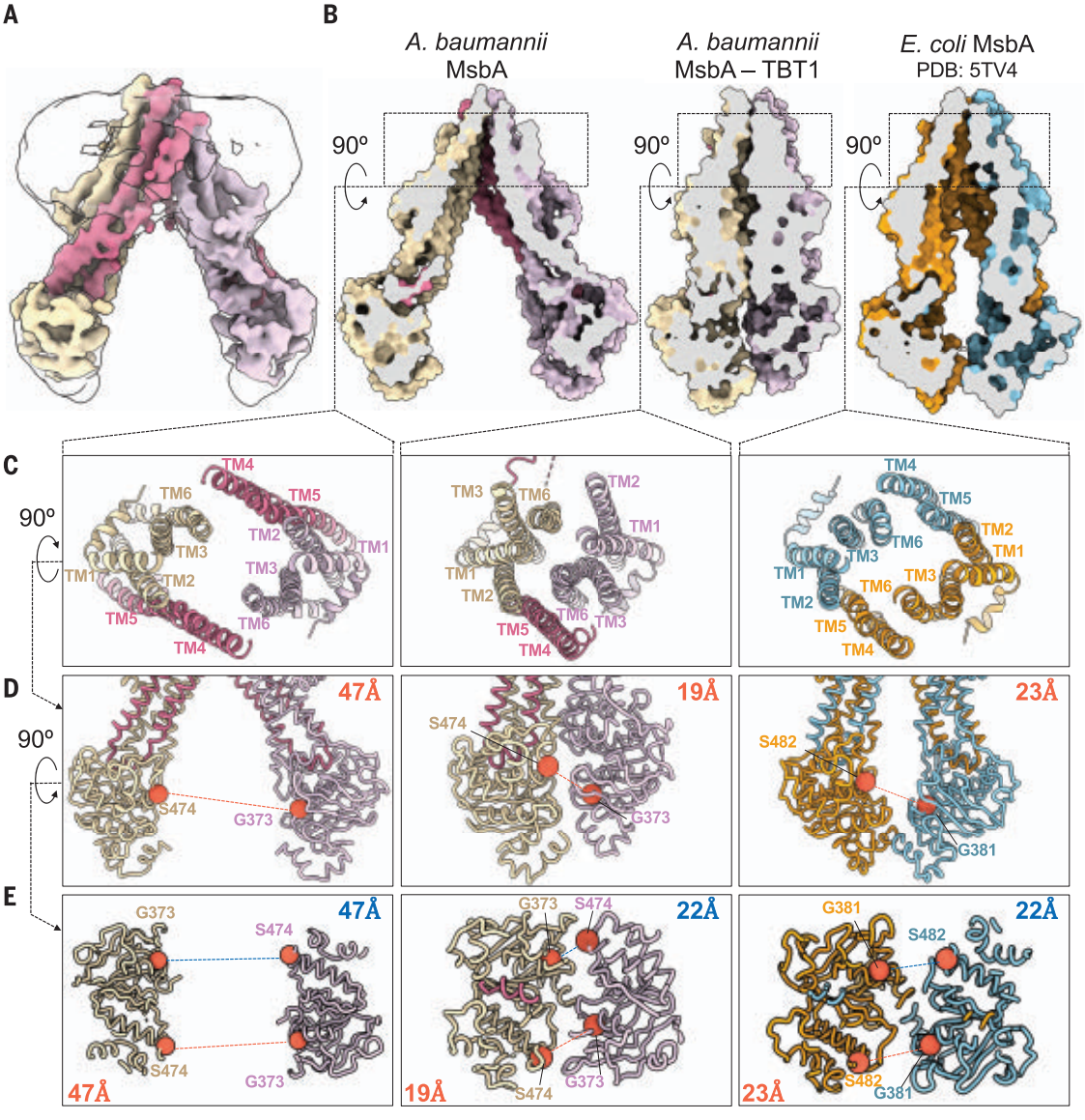


Fig. 2. Drug-free *A. baumannii* MsbA in nanodiscs adopts a wide inward-facing conformation. (**A**) Cryo-EM reconstruction (5.2-Å resolution) of *A. baumannii* MsbA in the absence of ligand. The unsharpened map filtered at 8-Å resolution is displayed as an outline to show the nanodisc. (**B**) Section through surface representation and comparison for three structures: *A. baumannii* MsbA, TBT1-bound *A. baumannii* MsbA, and *E. coli* MsbA (PDB ID: 5TV4) (*6*). (**C**) Top-down view of the substrate binding pockets. (**D**) Side views of the NBDs. Cα distances between ABC signature motif serine (S) and opposing Walker A glycine (G) residues are indicated as dashed lines, with values rounded to the nearest integer shown in each panel. (**E**) Top view of the NBDs, seen from the TMDs. Cα distances between NBDs are indicated as in (D).

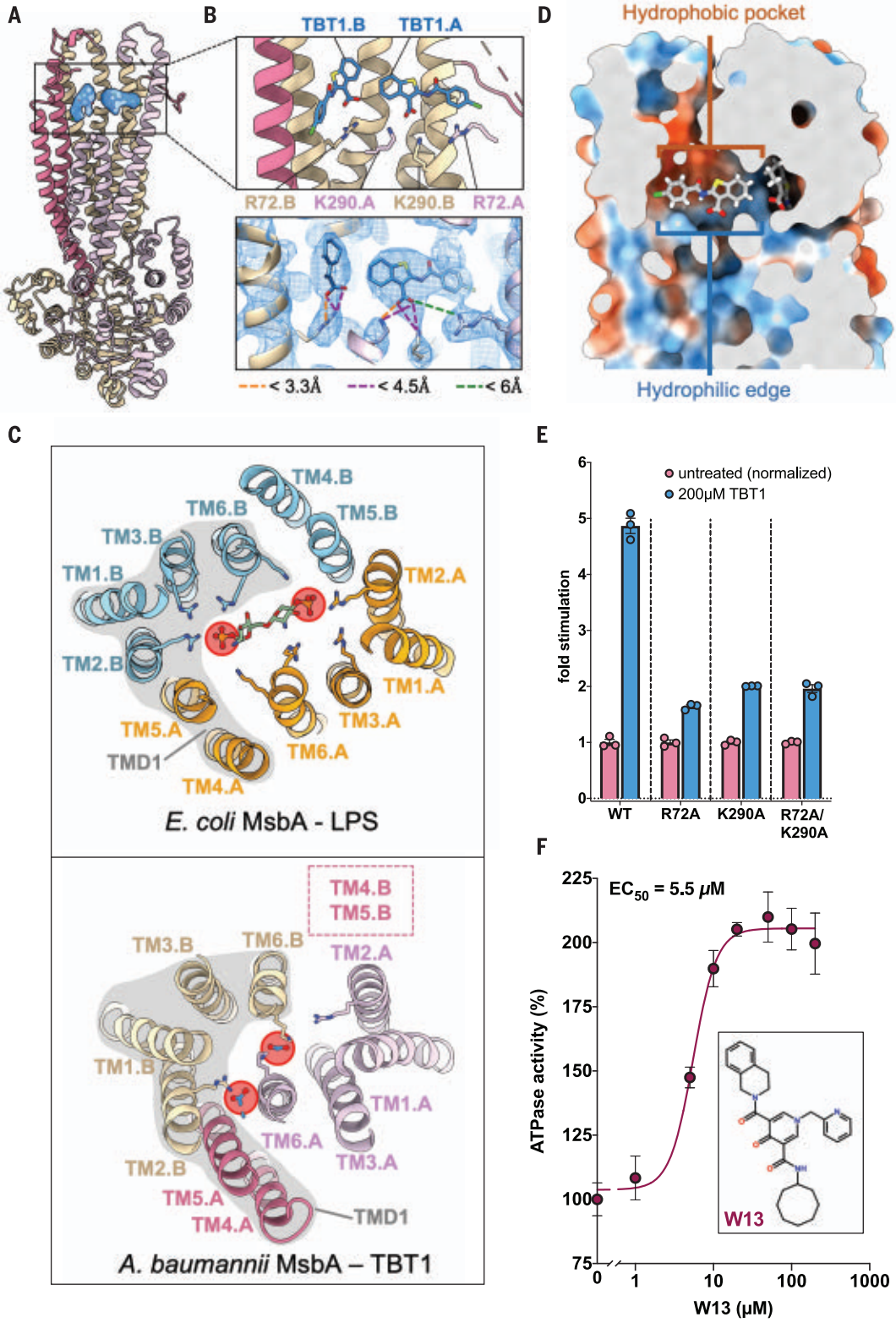


Fig. 3. TBT1 binding to MsbA. (**A**) Cartoon representation of TBT1-bound *A. baumannii* MsbA colored as in Fig. 1B, with TBT1 colored blue. (**B**) Close-up view of the TBT1 binding site (top) and superimposition of model and cryo-EM density (bottom). Distances between TBT1 carboxyl group and neighboring lysine (K) and arginine (R) residues are shown as colored dashed lines. (**C**) Top-down view of *E. coli* MsbA bound to LPS (PDB ID: 5TV4) (*6*) and *A. baumannii* MsbA bound to TBT1. TMD1 is conformationally similar in both structures and indicated with a gray background. The phosphate groups on LPS glucosamines and the carboxyl groups of TBT1 are marked with red circles. (**D**) Surface representation of the TBT1-binding pocket, with hydrophilic and hydrophobic surfaces colored blue and orange, respectively. (**E**) TBT1-induced ATPase stimulation of wild-type and mutant MsbA. The activity of each protein was normalized to its basal activity without TBT1. Error bars correspond to mean ± SD (from three independent replicate measurements). A, alanine; WT, wild type. (**F**) ATPase activity of *A. baumannii* MsbA with W13 at increasing concentrations. Each point represents mean ± SD (calculated from three independent data points). (Inset) The molecular structure of W13.

and unsharpened three-dimensional (3D) reconstructions (fig. S4B). This chain mismatch results in asymmetrical TMDs, with TMD1 including six helices (TM4.A, TM5.A, TM1.B, TM2.B, TM3.B, and TM6.B) and TMD2 including only four helices (TM1.A, TM2.A, TM3.A, and TM6.A). Whereas TM4.B is mostly unresolved, its N-terminal end is clearly discernable and forms a 61° angle with the C-terminal end of TM3.B, which is ~15° greater than the corresponding angle between the well-resolved TM4.A and TM3.A (fig. S4C). Accordingly, TM4.B kinks outward, which pushes the nanodisc density to bulge upward from the membrane plane (Fig. 1C and fig. S4C, left). Although the

consensus map features a bent nanodisc, a subset of particles was incorporated in comparatively flatter and larger nanodiscs yet still adopts the asymmetric TBT1-bound conformation, which suggests that TM4–TM5 bundle destabilization is not dependent on nanodisc size (fig. S2B). The domain-swapping TM4–TM5 bundle is of critical importance because the coupling helix (CH) at the TM4–TM5 junction (CH2), together with CH1 at the TM2–TM3 junction, is responsible for transmitting movement from the catalytically active NBDs to the TMDs (Fig. 1B, right). Because of the disordered TM4–TM5.B, CH2.B is absent for interaction with the NBD in chain A (NBD.A). Accordingly, NBD.A tilts up, moving toward the NBD dimerization interface (Fig. 1B). CH1.A also appears destabilized, likely because it cannot form proper interaction with the substantially shifted NBD.A (fig. S4D). The raised NBD.A is disengaged from its two CHs and seems decoupled from the TMDs.

At first sight, 2D class averages and 3D reconstruction of the cryo-EM images of TBT1-bound MsbA (Fig. 1D and fig. S4E) resemble those of nucleotide-bound MsbA (*6*), with both exhibiting constricted TMDs and reduced inter-NBD distance compared with those of nucleotide-free MsbA. However, the TBT1-bound MsbA conformation is still inward-facing because no nucleotide is present in the cryo-EM sample, and the NBDs remain fully separate. Accordingly, TMD1 is well aligned to the structure of nucleotide-free *E. coli* MsbA (root mean square deviation of 5.5 Å over 314 Cα atoms) (fig. S4F), which leaves substantial structural mismatch from TMD2. The most noteworthy structural rearrangement in TMD2 is the 13-Å displacement of TM6.A toward the central cavity (fig. S4G), which seems to pull NBD.A upward and into close proximity of NBD.B. Thus, the TBT1-bound MsbA structure described here is an unusual "collapsed inward-facing" conformation characterized by two previously unknown features: (i) the complete destabilization of one domain-swapping TM4–TM5 helix bundle and (ii) the highly asymmetric positioning of the NBDs.

TBT1 binding drives the closure of a wide-open MsbA

We then sought to characterize the structure of drug-free *A. baumannii* MsbA to understand how TBT1 binding affects MsbA conformation. To this end, we imaged *A. baumannii* MsbA in identical conditions as described above, but without the addition of TBT1. The resulting cryo-EM reconstruction at an overall resolution of 5.2 Å enabled unambiguous tracing of each TM helix, revealing a wide inward-facing conformation (Fig. 2A and fig. S5). To our knowledge, this is a previously uncharacterized structure of membrane-embedded MsbA in a wide inward-facing conformation. Such conformation had only been characterized with x-ray crystallography by using MsbA in detergents (*21*, *24*), and its physiological relevance has been heavily debated (*6*, *26*, *34*). Our result demonstrates that the wide inward-facing conformation of MsbA can exist in a lipid bilayer and may not be an artifact caused by deprivation of membrane environment and usage of detergents. Furthermore, the wide opening of MsbA seems not to be due to lack of bound substrate because a clear LPS-like density is present in the inner cavity (fig. S5G).

Because drug-free *A. baumannii* MsbA has a well-defined domain-swapping TM4–TM5 bundle in both TMDs and distanced NBDs, TBT1 binding is solely responsible for converting the wide-open MsbA to the collapsed inward-facing conformation. To gain insights into TBT1-stimulated ATP hydrolysis of MsbA, we compared three cryo-EM structures: TBT1-bound or drug-free *A. baumannii* MsbA and *E. coli* MsbA (Fig. 2, B to E). The central substrate-binding pocket of TBT1-bound MsbA is much more constricted than that in the other two structures (Fig. 2, B and C), which is consistent with the notion that the TMDs in TBT1-bound state present similarities with an outward-facing transporter. Upon TBT1 binding, the NBD distance is substantially decreased from ~47 to ~20 Å and even slightly shorter than the NBD distance in LPS-bound *E. coli* MsbA (Fig. 2, D and E). The two NBDs of TBT1-bound MsbA are positioned asymmetrically so that one ATP-binding site is tightened compared with the opposite ATP site (~19 versus ~22 Å) (Fig. 2E). The structural analysis of TBT1-bound MsbA provides an intuitive explanation for ATPase activity stimulation: The removal of the TM4–TM5.B bundle and the sliding of TM6.A into the central cavity greatly reduce inter-NBD distance, which thus increases the speed of NBD dimerization and ATP hydrolysis.

TBT1 hijacks the LPS-binding site to modulate MsbA

Two densities, each consistent with a TBT1 molecule, are present in the upper region of the TMDs (Fig. 3A). Although the structure resolution is limited, the distinctive features of the TBT1 densities allow for reliable modeling of the compounds and some critical side chains (Fig. 3B and fig. S6, A to C). The two TBT1 ligands are positioned asymmetrically in the central substrate pocket of MsbA and related by a ~60° rotation angle, with the carboxyl group pointing downward to the cytoplasm. The general binding mode of TBT1 is reminiscent of how *E. coli* MsbA binds the amphiphilic LPS by using a large hydrophobic pocket, which accommodates the lipid acyl chains in LPS, and a ring of basic residues (Arg^{78}, Arg^{148}, and Lys^{299}), which stabilize the phosphate groups on glucosamines (Fig. 3C, top) (*6*). Similarly, the two TBT1 molecules are located right above the ring of basic residues, with the aromatic ring structures of TBT1 positioned in the hydrophobic pocket (Fig. 3D).

Despite being in asymmetric binding pockets (fig. S6B), each TBT1 molecule is stabilized by an analogous set of hydrogen-bonding and electrostatic interactions between TBT1 carboxyl groups and neighboring basic side chains. TBT1.A is oriented parallel to the membrane plane, with its carboxyl group within hydrogen-bonding distance of $Lys^{290.A}$ and forming long-range electrostatic interactions with $Lys^{290.B}$ and $Arg^{72.A}$, and TBT1.B is rotated relative to TBT1.A, with no causal relationship forming a salt bridge with $Arg^{72.B}$ (Fig. 3B and fig. S6A).

In our structure, the carboxyl group of TBT1 is analogous to the phosphate groups on LPS glucosamines, which both form electrostatic interaction with MsbA. We thus sought to mutate Arg^{72} and Lys^{290} to determine whether MsbA retains sensitivity to TBT1-induced ATPase stimulation. *A. baumannii* MsbA with R72A, K290A, or R72A and K290A mutation was purified as wild-type MsbA and reconstituted in nanodiscs (fig. S7A). All MsbA mutants demonstrated largely diminished TBT1-induced stimulation compared with the wild-type protein, which highlights the importance of electrostatic interactions in TBT1 binding (Fig. 3E and fig. S7, B and C). Consistently, chemical modification of the carboxyl group of TBT1 renders the compound incompetent for MsbA inhibition (*28*). Although the electrostatic interactions described here may be the main drivers of TBT1 binding, hydrophobic contacts through the ring structures of TBT1 also participate in stabilizing the inhibitor. For example, L68F and L150V mutants, which were previously shown to confer resistance to TBT1 (*28*), are in the vicinity of the TBT1-binding sites (fig. S7D). Although they are not directly adjacent to TBT1, these mutations might allosterically alter the pocket and prevent efficient inhibitor binding.

Despite the strong resemblance in LPS and TBT1 recognition by MsbA, differences in TM positioning lead to divergence in ligand binding (Fig. 3C). LPS and TBT1 stabilization both involve TM2 (Arg^{72}) and TM6 (Lys^{290}), although unlike LPS, TBT1 does not clearly interact with TM3. The two TBT1 molecules together form a distorted mimic of LPS, with decreased distance between TBT1 carboxyl groups compared with LPS glucosamine phosphates. Upon TBT1 binding, TMD2 (TM1.A, TM2.A, TM3.A, and TM6.A) undergoes substantial changes: TM6.A moves into the center of the inner cavity, which presents $Lys^{290.A}$ for interaction with the hydroxyl group of TBT1.A and drags the connected NBD.A toward the

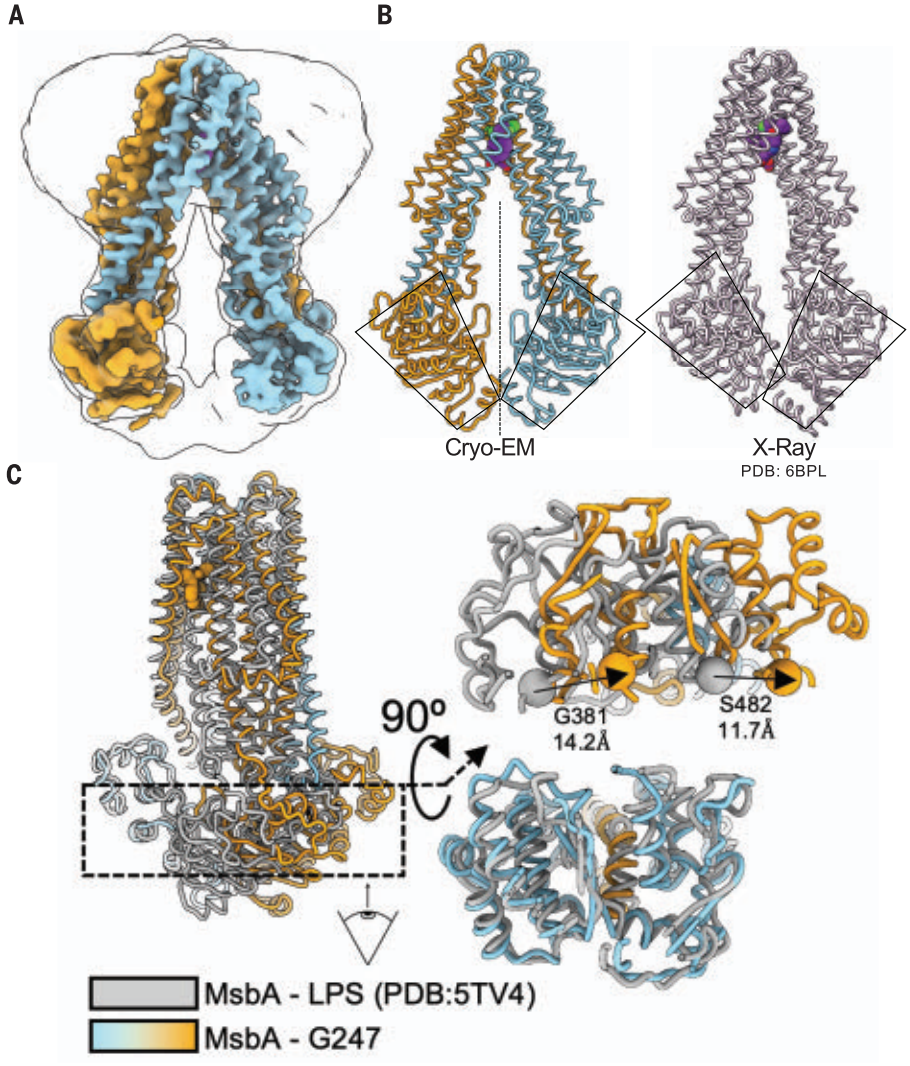


Fig. 4. G247 symmetrically increases inter-NBD spacing. (**A**) Cryo-EM reconstruction (3.9-Å resolution) of *E. coli* MsbA in complex with G247. The unsharpened map filtered at 6-Å resolution is displayed as an outline to show the nanodisc. (**B**) Comparison of MsbA structures determined by cryo-EM or x-ray crystallography (PDB ID: 6BPL) (*25*). The cryo-EM structure adopts C2 symmetry, whereas NBDs are asymmetrically positioned in the crystal structure. (**C**) Comparison of the cryo-EM structures of drug-free (PDB ID: 5TV4, gray) (*6*) and G247-bound (colored) MsbA, both in nanodiscs. G247 pushes NBDs away from each other, compared with the uninhibited state. The NBD shift is apparent when viewing MsbA from (left) the side and from (right) the cytoplasmic space. The dashed box in the MsbA side view (left) represents the cross section of the NBDs shown at right.

central axis (Fig. 3C, bottom). TMD2 then collapses around TM6.A, with TM2.A moving closer to TM6.B. The convergence between TM2.A and TM6.B might then push TM4–TM5.B laterally. The unsolvable contradiction between NBD.A moving toward the central axis and TM4–TM5.B being pinched outward results in the dissociation of CH2.B from NBD.A.

TBT1 offers an example of a small-molecule inhibitor that targets the central substrate-binding site of a nonmultidrug ABC transporter. Because most ABC transporters have relatively narrow substrate spectra, our findings suggest that mimicking substrate binding can be a generally applicable strategy for developing small-molecule modulators for all ABC transporters. To test whether the central pocket is useful for structure-based compound discovery, we performed an in silico chemical library screen against the TBT1 induced-fit pocket. Virtual screening of ~800,000 compounds, followed by ATPase assay of select hits, resulted in the identification of W13, a compound that stimulated ATPase activity with a median effective concentration lower than that of TBT1 (~5.5 versus 13 μM) (Fig. 3F and fig. S8B). The same MsbA mutants that were insensitive to TBT1 stimulation of ATPase activity also demonstrated lower W13-induced stimulation (fig. S7, B and C). W13 is structurally dissimilar to TBT1 and considerably larger in size (499 versus 336 Da) (fig. S8A), and a single W13 molecule is expected to occupy the central MsbA pocket (fig. S8, C and D). Molecular dynamics simulation of W13-bound MsbA in a POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) lipid bilayer suggests a binding mode that is reminiscent of TBT1, including hydrophobic groups of W13 extending upward in the hydrophobic pocket and carbonyl groups of W13 near TBT1-recognizing arginine and lysine residues (fig. S8D).

G247 symmetrically increases inter-NBD spacing and prevents MsbA closure

As the only other class of known MsbA-specific inhibitors, G compounds are mechanistically distinct from TBT1 because they bind the inward-facing MsbA and prevent ATP hydrolysis (*25*). The crystal structures of G compound–bound *E. coli* MsbA in facial amphiphile–3 (FA-3) detergent present asymmetrical NBDs positioning (*25*), although how the asymmetric conformation translates to impaired conformational cycling remains unclear. Furthermore, the extraordinary conformational flexibility of ABC transporters can result in different states being captured, depending on whether the protein is studied with crystallography (*21*, *22*, *24*) or cryo-EM (*6*, *26*). We thus sought to further investigate the mechanism of G compounds and acquired cryo-EM datasets of G247-bound *E. coli* MsbA in nanodisc (Fig. 4A and fig. S9) and DDM (fig. S10).

In previous crystal structures of G compound–bound MsbA (*25*), the NBDs are brought in close proximity and positioned asymmetrically, with one NBD raised relative to the other (Fig. 4B, right). By contrast, our cryo-EM structures in nanodisc and detergent are essentially identical; both exhibit clear C2 symmetry, even when refined without symmetry constraints (Fig. 4B, left, and fig. S11, B and C). Asymmetry in the crystal structures stems from Arg^{190}, which forms a salt bridge with the acrylic acid tail of the G compounds only in one subunit of MsbA (*25*). In cryo-EM structures, Arg^{190} is best ordered in the higher-resolution DDM structure and appears too distant from G247 to form a salt bridge (fig. S11E). Furthermore, the TM4, TM5, and TM6 bundle is positioned differently relative to TM1, TM2, and TM3 in the crystal and cryo-EM structures (fig. S11F), which ultimately affects the spacing between CHs (CH1 and CH2) and NBD positioning. Whereas CH1–CH2 spacing is reduced in crystal structures

compared with that of the drug-free conformation, CH1–CH2 distance is symmetrically increased in the cryo-EM structures (fig. S11G). Because the closed asymmetric conformation in crystal structures is not observed in any round of 3D classification (figs. S9E and S10A) or *cryoSPARC* (Structura Biotechnology, Canada) variability analysis (fig. S12A) (*35*) of cryo-EM images, we posit that the open C2-symmetric state is the predominant conformation of G247-bound MsbA in solution.

Comparing the cryo-EM structures of G247-bound and drug-free MsbA (*6*) reveals a mechanism of G compound inhibition that is different from previously anticipated: Instead of pushing one NBD toward the dimerization interface by 10 to 15 Å (*25*), G247 binding displaces the NBDs symmetrically and away from each other, which increases the inter-NBD distance by ~13 Å (Fig. 4C). This model is consistent with 3D variability analysis (fig. S12A), in which G247-bound MsbA NBDs move away from the dimerization axis relative to the NBDs of drug-free MsbA. Whereas drug-free MsbA has NBDs aligned in a head-to-tail manner and thus primed for ATP hydrolysis, the increased inter-NBD distance in G247-bound MsbA is expected to reduce NBD dimerization efficiency and ATPase activity. To formally test this hypothesis, we exposed MsbA to the nonhydrolyzable nucleotide AMP-PNP in the presence or absence of G247 (fig. S12B). When subjected to 1 mM AMP-PNP, nearly all MsbA particles showed a closed conformation with dimerized NBDs. By contrast, 5 μM of G247 was sufficient to shift >60% of particles to an open conformation with separate NBDs (fig. S12B).

Discussion

The effect on MsbA conformation and the inhibition mechanism of TBT1 (fig. S13) are different from those of small-molecule modulators for other ABC transporters, which typically induce more modest conformational changes (for example, ABCB1 and ABCG2 in fig. S6D). The narrowed inter-NBD spacing of TBT1-bound MsbA is more akin to the NBD tightening that is observed when the human multidrug transporter ABCC1 is exposed to its endogenous leukotriene substrate (fig. S6D) (*1*), which further suggests that TBT1 triggers substrate recognition–like behavior in MsbA by mimicking the much larger LPS substrate. Yet, unlike the authentic ligand, TBT1 breaks the symmetry of homodimeric MsbA by destabilizing one domain-swapping helix bundle. Uncoupling of ATPase function and substrate transport by TBT1 occurs through the domain-swapped CH, which is a landmark feature of the type IV exporter fold (*33*). Thus, small-molecule decouplers acting like TBT1 may be identified for many other type IV exporters such as ABCB1 and ABCC1. We have shown that G247, in contrast to TBT1, suppresses ATPase activity by increasing inter-NBD distance (fig. S13). This inhibition mechanism is similar to that of previously described ABC transporter inhibitors, including zosuquidar, which reduces the activity of the homologous ABCB1 transporter by pushing its NBDs away from each other (*17*). Because most ABC transporter inhibitors characterized so far are ATPase suppressors, it is likely that more compounds act similarly to G247 than to TBT1.

TBT1 and G247 currently face severe drawbacks that prevent their application in clinical settings, including low binding affinity to MsbA (*28*) and interaction with animal serum (*29*). Yet compound-bound ABC transporter structures provide essential templates for structure-based drug development because molecules such as TBT1 and G247 reveal unexpected binding pockets in previously unidentified MsbA conformations. Our identification of W13 as an ATPase stimulator would have been unlikely if the structure of TBT1-bound MsbA were not available. One promising direction for compound optimization may consist of tethering together two TBT1-like molecules to take advantage of cooperative binding. Although TBT1 only works on *A. baumannii* MsbA, future efforts may yield molecules that are able to extend this inhibition mechanism to MsbA in other Gram-negative bacteria, including additional antibiotic-resistant pathogens.

REFERENCES AND NOTES

1. Z. L. Johnson, J. Chen, *Cell* **168**, 1075–1085.e9 (2017).
2. N. M. I. Taylor *et al*., *Nature* **546**, 504–509 (2017).
3. K. Nosol *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26245–26253 (2020).
4. Z. Zhang, J. Chen, *Cell* **167**, 1586–1597.e9 (2016).
5. N. Li *et al*., *Cell* **168**, 101–110.e10 (2017).
6. W. Mi *et al*., *Nature* **549**, 233–237 (2017).
7. H. Qian *et al*., *Cell* **169**, 1228–1239.e10 (2017).
8. J. A. Olsen, A. Alam, J. Kowal, B. Stieger, K. P. Locher, *Nat. Struct. Mol. Biol.* **27**, 62–70 (2020).
9. D. C. Rees, E. Johnson, O. Lewinson, *Nat. Rev. Mol. Cell Biol.* **10**, 218–227 (2009).
10. K. P. Locher, *Nat. Struct. Mol. Biol.* **23**, 487–493 (2016).
11. C. Thomas, R. Tampé, *Curr. Opin. Struct. Biol.* **51**, 116–128 (2018).
12. S. Srikant, R. Gaudet, *Nat. Struct. Mol. Biol.* **26**, 792–801 (2019).
13. D. Szöllősi, D. Rose-Sperling, U. A. Hellmich, T. Stockner, *Biochim. Biophys. Acta Biomembr.* **1860**, 818–832 (2018).
14. S. Shukla, S. Ohnuma, S. V. Ambudkar, *Curr. Drug Targets* **12**, 621–630 (2011).
15. A. M. Plummer, A. T. Culbertson, M. Liao, *Annu. Rev. Physiol.* **83**, 153–181 (2021).
16. V. G. Lewis, M. P. Ween, C. A. McDevitt, *Protoplasma* **249**, 919–942 (2012).
17. A. Alam, J. Kowal, E. Broude, I. Roninson, K. P. Locher, *Science* **363**, 753–756 (2019).
18. A. Alam *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1973–E1982 (2018).
19. S. M. Jackson *et al*., *Nat. Struct. Mol. Biol.* **25**, 333–340 (2018).
20. B. J. Orlando, M. Liao, *Nat. Commun.* **11**, 2264 (2020).
21. C. F. Higgins, K. J. Linton, *Science* **293**, 1782–1784 (2001).
22. A. Ward, C. L. Reyes, J. Yu, C. B. Roth, G. Chang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19005–19010 (2007).
23. W. T. Doerrler, M. C. Reedy, C. R. Raetz, *J. Biol. Chem.* **276**, 11461–11464 (2001).
24. P. S. Padayatti *et al*., *Structure* **27**, 1114–1123.e3 (2019).
25. H. Ho *et al*., *Nature* **557**, 196–201 (2018).
26. G. Angiulli *et al*., *eLife* **9**, e53530 (2020).
27. F. Thélot, B. J. Orlando, Y. Li, M. Liao, *Curr. Opin. Struct. Biol.* **63**, 26–33 (2020).
28. G. Zhang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6834–6839 (2018).
29. M. K. Alexander *et al*., *Antimicrob. Agents Chemother.* **62**, e01142-18 (2018).
30. H. W. Boucher *et al*., *Clin. Infect. Dis.* **48**, 1–12 (2009).
31. I. G. Denisov, S. G. Sligar, *Chem. Rev.* **117**, 4669–4713 (2017).
32. M. A. McLean, M. C. Gregory, S. G. Sligar, *Annu. Rev. Biophys.* **47**, 107–124 (2018).
33. C. Thomas *et al*., *FEBS Lett.* **594**, 3767–3775 (2020).
34. J. Wright, S. P. Muench, A. Goldman, A. Baker, *Biochem. Soc. Trans.* **46**, 1475–1484 (2018).
35. A. Punjani, D. J. Fleet, *J. Struct. Biol.* **213**, 107702 (2021).

ACKNOWLEDGMENTS

We are grateful to D. Kahne and D. Dates for providing the initial TBT1 sample, as well as to Genentech for their donation of G247. We thank J. Zhang for helping with ATPase screening experiments. We thank Z. Li, S. Sterling, R. Walsh, and S. Rawson from the Harvard Cryo-EM Center for Structural Biology for their training and expert advice. The *A. baumannii* MsbA data and one dataset of *E. coli* MsbA-nanodiscs with G247 were collected at the cryo-EM core facility at University of Massachusetts Medical School, and the data of MsbA-DDM with G247 were acquired at the Harvard Cryo-EM Center for Structural Biology at Harvard Medical School. We are grateful to all Liao laboratory members for their helpful feedback throughout this project. **Funding:** M.L. was supported by an NIH grant from the NIGMS (R01 GM122797), and J.H. was funded by the National Natural Science Foundation of China (grant nos. 21803057 and 32171247). **Author contributions:** M.L. conceived and supervised the project. F.A.T. performed molecular cloning, protein purification, nanodisc reconstitution, ATPase assays, negative-stain EM, sample vitrification, cryo-EM data collection and processing, and model building. K.S. and C.X. helped with cryo-EM data acquisition. W.Z and J.H. performed virtual screen and molecular dynamics simulation. F.A.T. and M.L. drafted the manuscript. All authors contributed to data analysis and manuscript preparation. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Atomic models are available through the Protein Data Bank (PDB) under the IDs 7MET (TBT1-bound *A. baumannii* MsbA), 7RIT (drug-free *A. baumannii* MsbA), and 7MEW (G247-bound *E. coli* MsbA in nanodisc). Cryo-EM maps are available through the Electron Microscopy Data Bank under the accession codes EMD-23803 (TBT1-bound *A. baumannii* MsbA), EMD-23804 (drug-free *A. baumannii* MsbA), EMD-23805 (G247-bound *E. coli* MsbA in nanodisc), EMD-24446 (G247-bound *E. coli* MsbA in DDM, C2), and EMD-24447 (G247-bound *E. coli* MsbA in DDM, C1). Unprocessed motion-corrected micrographs for all datasets are accessible through EMPIAR under the accession codes 10778 (TBT1-bound *A. baumannii* MsbA), 10801 (drug-free *A. baumannii* MsbA), 10799 (G247-bound *E. coli* MsbA in nanodisc), and 10800 (G247-bound *E. coli* MsbA in DDM). All other data are present in the main text or supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi9009
Materials and Methods
Figs. S1 to S13
Tables S1 and S2
References (*36*–*57*)
MDAR Reproducibility Checklist

11 April 2021; accepted 14 September 2021
Published online 23 September 2021
10.1126/science.abi9009

NUCLEAR ARCHITECTURE

Single-cell nuclear architecture across cell types in the mouse brain

Yodai Takei[1], Shiwei Zheng[2], Jina Yun[1], Sheel Shah[1], Nico Pierson[1], Jonathan White[1], Simone Schindler[1], Carsten H. Tischbirek[1], Guo-Cheng Yuan[2], Long Cai[1]*

Diverse cell types in tissues have distinct gene expression programs, chromatin states, and nuclear architectures. To correlate such multimodal information across thousands of single cells in mouse brain tissue sections, we use integrated spatial genomics, imaging thousands of genomic loci along with RNAs and epigenetic markers simultaneously in individual cells. We reveal that cell type–specific association and scaffolding of DNA loci around nuclear bodies organize the nuclear architecture and correlate with differential expression levels in different cell types. At the submegabase level, active and inactive X chromosomes access similar domain structures in single cells despite distinct epigenetic and expression states. This work represents a major step forward in linking single-cell three-dimensional nuclear architecture, gene expression, and epigenetic modifications in a native tissue context.

A key feature of modern biology is enabling measurement of many parameters in the same cells to arrive at accurate low-dimensional representations of biological systems (*1*, *2*). In the case of three-dimensional (3D) organization of the genome, which is critical for many cellular processes, from regulating gene expression to establishing cellular identity (*3*), it is crucial to integrate chromosome structure measurements (*4*, *5*) with others that capture transcriptional states (*6*, *7*), chromatin states (*8*), and nuclear bodies (*9*, *10*). However, multimodal measurements of chromosome structure, transcriptional states, and epigenetic markers in the same cell remain challenging with sequencing approaches (*2*). Imaging-based approaches, on the other hand, allow direct integration of multimodal measurements, including chromosome structures (*6*, *7*, *11*–*17*). We recently established an integrated spatial genomics approach (*17*), imaging thousands of genomic loci along with transcriptional states and subnuclear localization of histone modifications and nuclear bodies in single mouse embryonic stem (ES) cells. This approach revealed deterministic scaffolding of chromosomes across multiple nuclear bodies and protein globules in individual cells (*17*). Nevertheless, it is unknown to what extent principles for single-cell nuclear organization extend to diverse cell types in complex mammalian tissues.

Integrated spatial genomics in the brain

To investigate single-cell nuclear architecture across cell types, we imaged sections of the adult mouse cerebral cortex with 3660 DNA loci, 76 cellular RNAs, and 8 chromatin marks and nuclear bodies over 125 rounds of hybridizations and imaging while maintaining tissue integrity (see methods in the supplementary materials) (*18*) (Fig. 1, A and B; figs. S1 to S4; and tables S1 to S4). Specifically, we used DNA sequential fluorescence in situ hybridization (DNA seqFISH+) to image 2460 loci at ~1-Mb resolution across the genome and an additional 1200 loci for at least a 1.5-Mb region in each chromosome at 25-kb resolution (*17*) (Fig. 1, C to E; fig. S1, B and C; and table S1). We detected 2813.0 ± 1334.0 (median ± SD) spots per cell in total with low false positives in 2762 cells from three biological replicates (Fig. 1, F and G, and fig. S2, A and H), corresponding to an estimated detection efficiency of at least 38.4 ± 18.2% (median ± SD) (methods). The DNA seqFISH+ data agreed well with sequencing-based bulk Hi-C data (*19*) and among the three biological replicates (Fig. 1H and fig. S3), demonstrating the robustness of DNA seqFISH+ for mapping 3D chromosome structures in tissue samples.

We clustered the RNA seqFISH data and obtained nine major cell type clusters within the cerebral cortex that matched well with cell types identified from single-cell RNA sequencing (*20*), including excitatory neurons; four subclasses of inhibitory neurons expressing parvalbumin (Pvalb), vasoactive intestinal polypeptide (Vip), neuron-derived neurotrophic factor (Ndnf), or somatostatin (Sst); three types of glial cells (astrocytes, microglia and oligodendrocyte precursor cells, and oligodendrocytes); and endothelial cells (Fig. 1I and fig. S1, C to F).

In addition to the genome and RNA imaging, we used sequential immunofluorescence (IF) with oligonucleotide-conjugated primary antibodies (methods) (*17*) to detect six histone modifications or variants (H3K4me2, H3K27me2, H3K27me3, H3K9me3, H4K20me3, and mH2A1), nuclear speckles [splicing factor 3A subunit 2 (SF3a66)], and the methyl-CpG-binding protein MeCP2 (Fig. 2A, figs. S4 and S5A, and table S3). Additionally, we measured three noncoding RNAs that mark the inactive X chromosome (Xi, Xist), the nucleolus [internal transcribed spacer region 1 (ITS1)], nuclear speckles [metastasis-associated lung adenocarcinoma transcript 1 (Malat1)], as well as five repetitive regions [long interspersed element 1 (LINE1), short interspersed nuclear element B1 (SINEB1), major satellite (MajSat), minor satellite (MinSat), and Telomere] (*21*, *22*) (figs. S4 and S5A). The localization patterns of individual antibodies, RNA, and DNA markers in our data were consistent with literature (*21*, *22*) (fig. S4), enabling the multimodal investigation of single-cell nuclear architecture across cell types with unprecedented detail.

Distinct nuclear features across cell types in the brain

The nine major cell types displayed clear differences in the global levels of both repressive marks (MeCP2, H3K27me3, and mH2A1) and an active mark (H3K4me2) (Fig. 2A and figs. S5 and S6). Clustering of single cells using the multiplexed IF data was able to distinguish the same nine cell types identified by RNA seqFISH (Fig. 2B), supporting a strong correlation between global chromatin states and transcriptional states. Notably, even the 4′,6-diamidino-2-phenylindole (DAPI) features alone were sufficient to separate the major cell types in the cortical areas of the mouse brain (Fig. 2, A, C, and D, and fig. S5C).

The spatial versus genomic distance scaling was distinct in different cell types in the 1-Mb-resolution data (Fig. 2, E and F, and figs. S7 and S8), which cannot be explained just by nuclear size differences (fig. S5C). In contrast, in the 25-kb-resolution data at the targeted megabase regions, the scaling relationship differed depending on the chromosomal regions and cell types (Fig. 2, E and F, and figs. S7 and S8). To obtain a more integrative picture of nuclear organization, we need to examine the association between DNA and nuclear bodies beyond characterizing individual components.

Cell type–specific chromatin profiles can relate to differential transcriptional states

To characterize the spatial association between DNA loci, chromatin marks, and nuclear bodies, we calculated the fraction of instances in which each DNA locus is detected within 300 nm from the exterior of each mark (*17*, *23*) (Fig. 3, A to D, and figs. S9 to S12). With these imaging-based "chromatin profiles" using super-resolved DNA seqFISH+ spots and diffraction-limited IF images, we have previously showed a high correlation with sequencing-based bulk measurements such

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [2]Department of Genetics and Genomic Sciences and Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
*Corresponding author. Email: lcai@caltech.edu

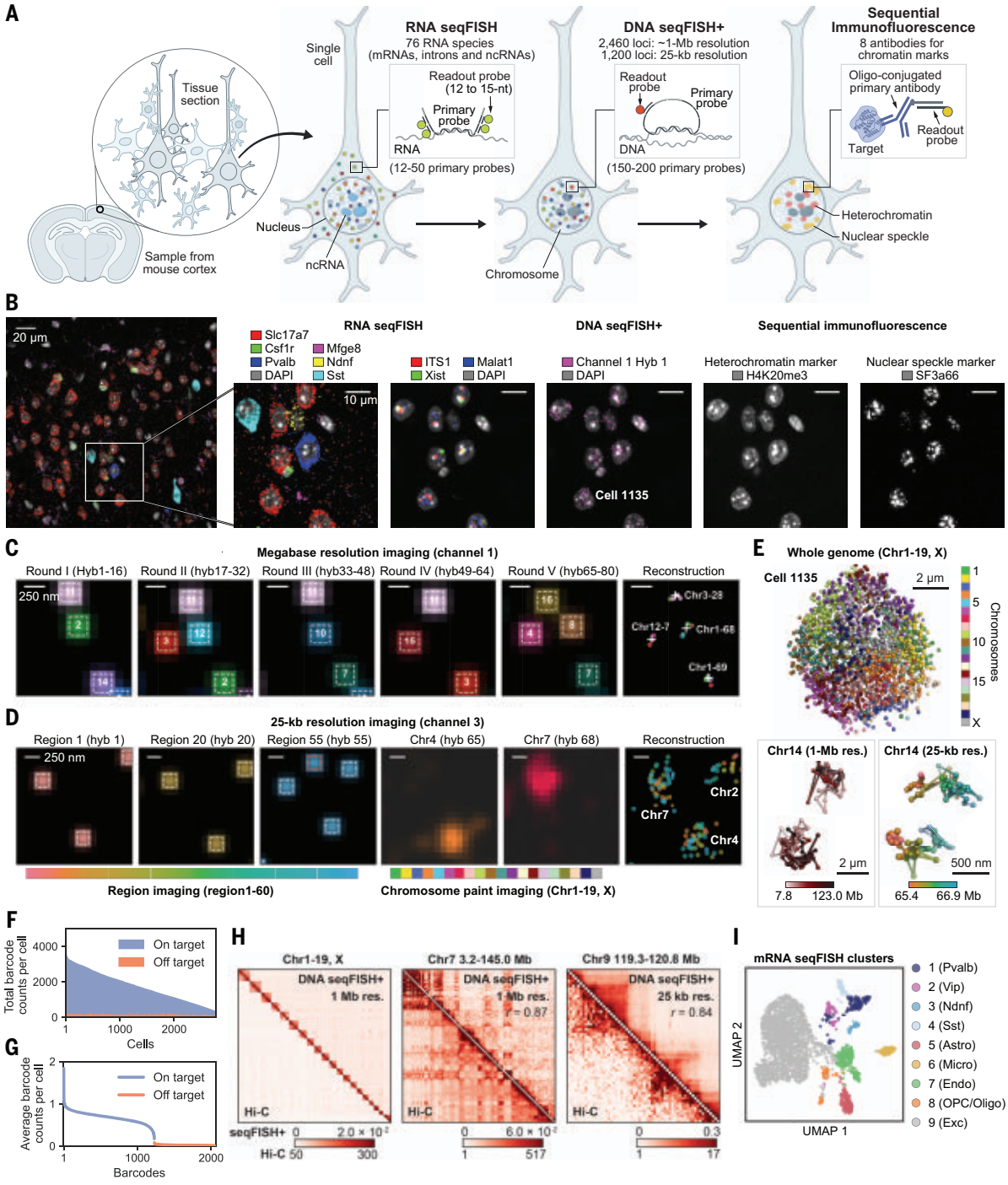


Fig. 1. Integrated spatial genomics in the mouse brain. (**A**) Schematic of integrated RNA seqFISH, DNA seqFISH+, and sequential immunofluorescence (IF) measurements in the mouse brain cortex. ncRNA, noncoding RNA; nt, nucleotides. (**B**) Example RNA seqFISH, DNA seqFISH+, and sequential IF images in a brain slice. The images are the maximum intensity z-projected from 1.5-μm z-slices. (**C** and **D**) The zoomed-in view of cell 1135 for 1-Mb resolution DNA seqFISH+ through five rounds of barcoding with 16 pseudocolors from a single z-section in (C) and for 25-kb resolution DNA seqFISH+ through 60 hybridization rounds of adjacent region imaging followed by 20 hybridization rounds of chromosome painting shown with pseudocolors with spots from all z-slices in (D). White boxes on pseudocolor spots indicate identified barcodes, and the red box indicates a rejected nonspecific binding spot. (**E**) 3D reconstruction of a single nucleus colored by chromosomes (top) or chromosome coordinates (bottom). (**F**) Frequencies of on- and off-target barcodes in fluorescent channels 1 and 2 per cell. (**G**) Average frequencies of individual on-target and off-target barcodes in (F). (**H**) Agreement between the spatial proximity maps (probability of pairs of loci within 500 nm in cells for 1-Mb-resolution data and within 150 nm in alleles for 25-kb-resolution data) from DNA seqFISH+ (upper right) and normalized Hi-C read counts (lower left) at different genomic scales. Hi-C data are displayed with 1-Mb and 25-kb bin sizes to compare with 1-Mb and 25-kb resolution DNA seqFISH+ data, respectively. Hi-C data from mouse brain cortex originally at 40-kb binning (*19*) was recomputed at 25-kb binning for direct comparison. (**I**) Uniform manifold approximation and projection (UMAP) representation of the cell type clusters determined by mRNA seqFISH profiles. n = 2762 cells from three biological replicates in (F) to (I). Astro, astrocytes; Micro, microglia; Endo, endothelial cells; OPC/Oligo, oligodendrocyte precursor cells and oligodendrocytes; Exc, excitatory neurons.

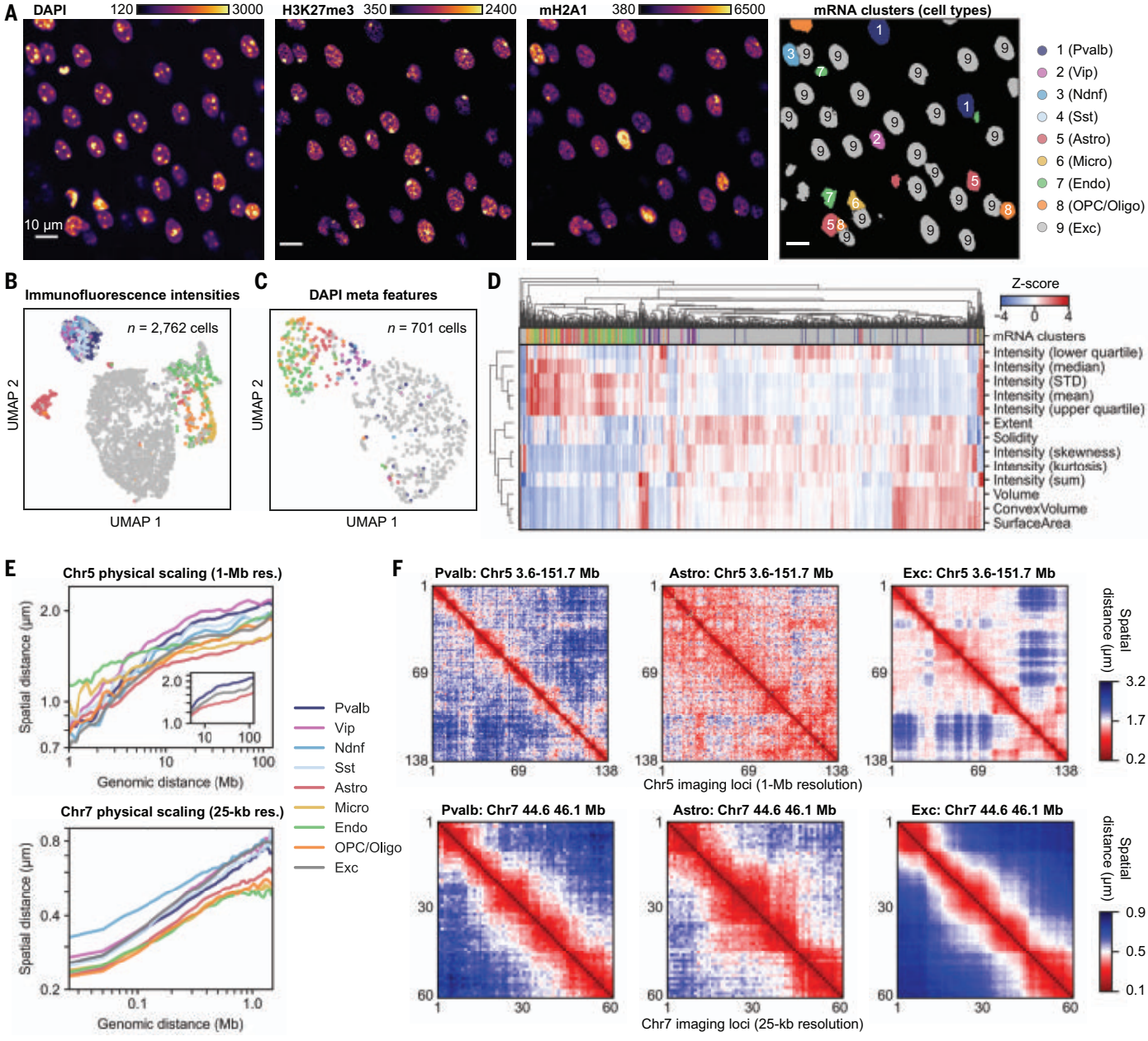


Fig. 2. Nuclear morphology, global chromatin states, and chromosome scaling are cell type dependent. (**A**) DAPI, H3K27me3, and mH2A1 staining from a single z-section superimposed on the transcriptionally defined cell clusters. Color bars represent fluorescence intensity (a.u., arbitrary units). (**B**) UMAP representation shows separation of excitatory, inhibitory, and glial cells on the basis of overall intensities for eight IF markers in single cells. A total of 2762 cells from three biological replicates were used. (**C**) UMAP representation of cells based on DAPI meta features used in (D). Cells are colored by their transcriptional cell types in (B) and (C). (**D**) Hierarchical clustering of the DAPI meta features compared with the RNA seqFISH clusters shown on top. A total of 701 cells in the center z-sections from three biological replicates were used in (C) and (D). (**E**) Physical distance as a function of genomic distance across transcriptional defined cell types. (**F**) Spatial distance between pairs of intrachromosomal loci within cells by DNA seqFISH+ in different cell types at 1-Mb resolution (top row; quartile spatial distance within cells) and at 25-kb resolution (bottom row; median spatial distance within alleles). A total of 2762 cells were used for (E) and 155 cells for Pvalb, 152 cells for astrocytes, and 1895 cells for excitatory neurons were used in (F) from three biological replicates.

as chromatin immunoprecipitation sequencing (ChIP-seq) and split-pool recognition of interactions by tag extension (SPRITE) at 1-Mb resolution in mouse ES cells (*17*).

Some chromatin profiles showed specific patterns that varied between cell types at 1-Mb resolution (Fig. 3, A to D, and figs. S9 to S12). The chromatin profiles for constitutive heterochromatin stained by DAPI were distinct between neurons and astrocytes (Fig. 3A and fig. S10, A and B), even though the associations were typically enriched in the centromere-proximal genomic loci (*9*, *17*) and adenine-thymine (AT)–rich regions (*24*) in all cell types (Fig. 3A and fig. S10C). Furthermore, the chromatin profiles for H3K27me3 were distinct across cell types in the brain and mouse ES cells (Fig. 3B and fig. S11).

DNA loci associated with nuclear speckle markers (SF3a66 and Malat1) were highly correlated among different cell types and with

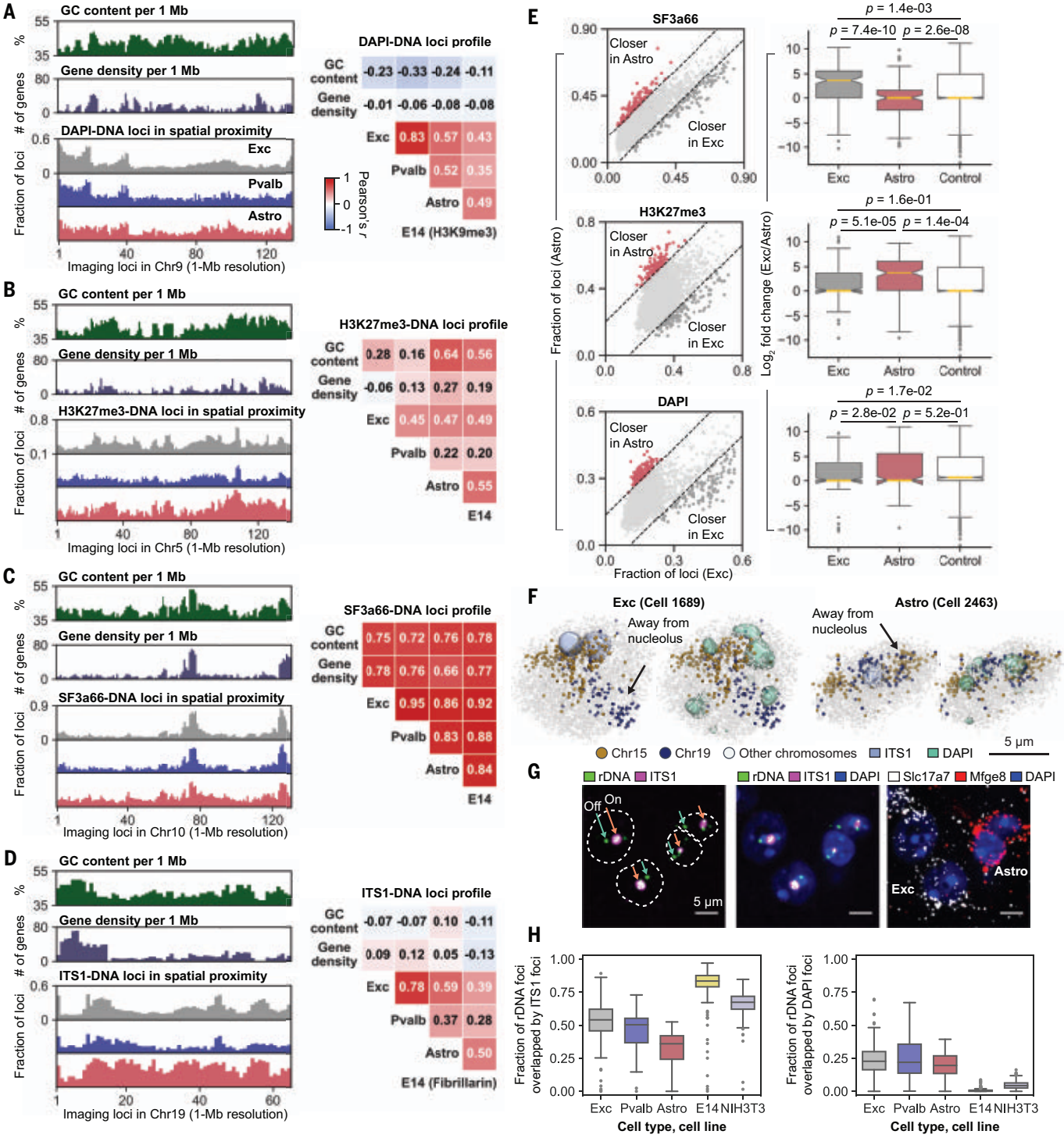


Fig. 3. Chromatin profiles and proximal points across cell types. (**A** to **D**) GC content, gene density, and cell type–specific chromatin profiles defined by the fraction of loci found within 300 nm of DAPI, H3K27me3, SF3a66, and ITS1 exteriors (left panels). Pearson correlation between those (n = 2340 autosomal loci; right panels). (**E**) Comparison of fraction of loci for each marker between excitatory neurons and astrocytes (n = 2340 autosomal loci) (left panels) and fold change of differentially expressed genes between cell types at top 5% differentially associated loci to each mark colored in left panels (n = 117 loci for each cell type) or other loci as control (n = 2106 loci) (right panels). P values by two-sided Wilcoxon's signed rank-sum test. n = 1895, 155, and 152 cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes, respectively, from three biological replicates in (A) to (E). (**F**) Representative 3D images of nucleoli by ITS1 RNA FISH, heterochromatic bodies stained by DAPI, and chromosomes 15 and 19 by DNA seqFISH+ in the excitatory neuron (left) and astrocyte (right). (**G**) rDNA staining confirms that some of the ribosomal repeat regions appear outside of the nucleolus and are located at DAPI-enriched heterochromatin in postmitotic cells in the brain. The images are from a single z-section. (**H**) The fraction of rDNA loci present in the nucleolus or close to heterochromatic regions. n = 365, 18, and 32 for excitatory neurons, Pvalb inhibitory neurons, and astrocytes, respectively, from five fields of view in one mouse brain section, and n = 305 and 334 cells for E14 mouse ES cells and NIH3T3 cells from seven and eight fields of view in each sample.

gene densities and guanine-cytosine (GC) content, at 1-Mb resolution (Fig. 3C and figs. S9A and S12). However, some DNA loci are differentially associated with nuclear speckles at 1-Mb resolution between cell types. These small fraction changes toward nuclear speckles are often associated with transcriptional upregulation of genes in those regions (Fig. 3E and fig. S13). Conversely, transcriptionally upregulated loci between cell types increased association toward nuclear speckles (fig. S14). In comparison, we only observed transcriptional decrease at loci with increased H3K27me3 association or DAPI association in specific cell type comparisons (Fig. 3E and fig. S13), suggesting a more subtle contribution at this resolution. Our observation extends to tissues the findings by tyramide signal amplification sequencing (TSA-seq) in cell lines (*25*) and includes multiple chromatin markers.

We observed a similar cell type–dependent association of chromosomes with the nucleolus stained by internal transcribed spacer 1 (ITS1) noncoding RNA in the mouse brain cortex (Fig. 3D). We observed that the 45*S* ribosomal DNA (rDNA)–containing chromosomes 15 and 19 were not in physical proximity to the nucleoli but were close to DAPI-rich heterochromatin regions in some cells (Fig. 3F and fig. S15, A to C), possibly owing to rDNA silencing (*26*). To confirm this, we performed imaging of rDNA loci by DNA FISH and found that different fractions of rDNA loci were associated with the nucleolus and DAPI-rich heterochromatin regions among neurons, astrocytes, mouse ES cells, and cultured fibroblasts (Fig. 3, G and H, and fig. S15D), lending support to the notion that nucleolar organizer regions can be stably silenced (*26*) in a tissue-specific fashion.

Cell type–specific proximal points anchor chromosomes to nuclear bodies in single cells

We defined DNA loci that were consistently associated, at least 2 standard deviations above the mean association frequency, with a particular chromatin mark or nuclear body in each cell type as "proximal points" (methods) (fig. S16). For example, SF3a66 proximal points, such as *Nxph4* and *Malat1* loci, are within 300 nm of nuclear speckles at nearly 80% of the instances in single excitatory neurons (*23*). We previously observed that certain DNA loci for each IF marker consistently appear on the exterior of the respective marker in single mouse ES cells (*17*). This phenomenon extends to the mouse brain and appears to be consistent in different cell types, despite the differences in the morphological features and arrangement of nuclear bodies in individual neuronal and glial cells (fig. S16).

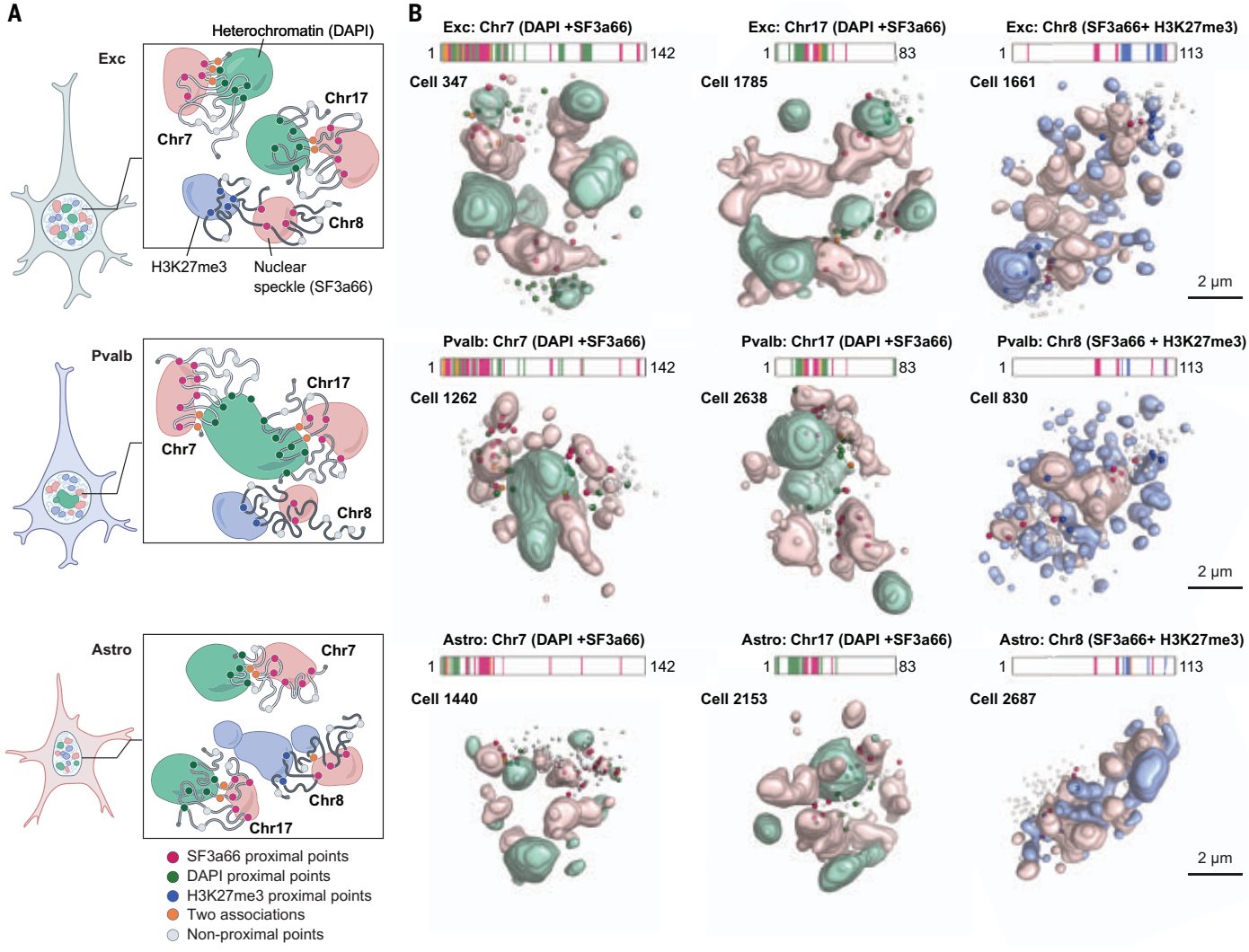


Fig. 4. Proximal points straddle across nuclear bodies. (**A**) Illustration showing chromosomes 7 and 17 with proximal points for SF3a66 and DAPI and chromosome 8 with proximal points for SF3a66 and H3K27me3 across cell types (Exc, excitatory neurons; Pvalb, Pvalb inhibitory neurons; Astro, astrocytes). (**B**) Representative 3D images of individual chromosomes and markers showing that each chromosome has cell type–specific proximal points and spans the corresponding nuclear bodies in single cells. The same coloring as (A).

Because proximal points have cell type–specific patterns for different chromatin marks across the genome, they can scaffold and constrain the nuclear organization of chromosomes by their association to the nuclear bodies or chromatin marks in individual nuclei (Fig. 4 and fig. S16). For example, chromosomes 7 and 17 have proximal points for nuclear speckles (SF3a66) and heterochromatic bodies (DAPI) in excitatory neurons, inhibitory neurons, and astrocytes, and both chromosomes straddled these nuclear bodies in single cells of all three cell types (Fig. 4B). Similarly, chromosome 8 has nuclear speckles and H3K27me3 proximal points in all three cell types and spans those nuclear globules in single cells (Fig. 4B). Together, these findings in the mouse brain show that chromatin proximal points serve as organizational invariants in the nuclei of single cells across cell types, despite the highly variable appearance of individual chromosomes and nuclear bodies in individual cells.

Cell type–specific nuclear bodies determine interchromosomal proximity and radial positioning

We examined whether the proximal point association to nuclear bodies in different cell types accounts for the cell type–specific interchromosomal interaction and radial chromosomal positioning (Fig. 5, A to D, and fig. S17). For example, chromosome 17 contains many heterochromatin proximal points in neurons (Fig. 4B). At the same time, a few large

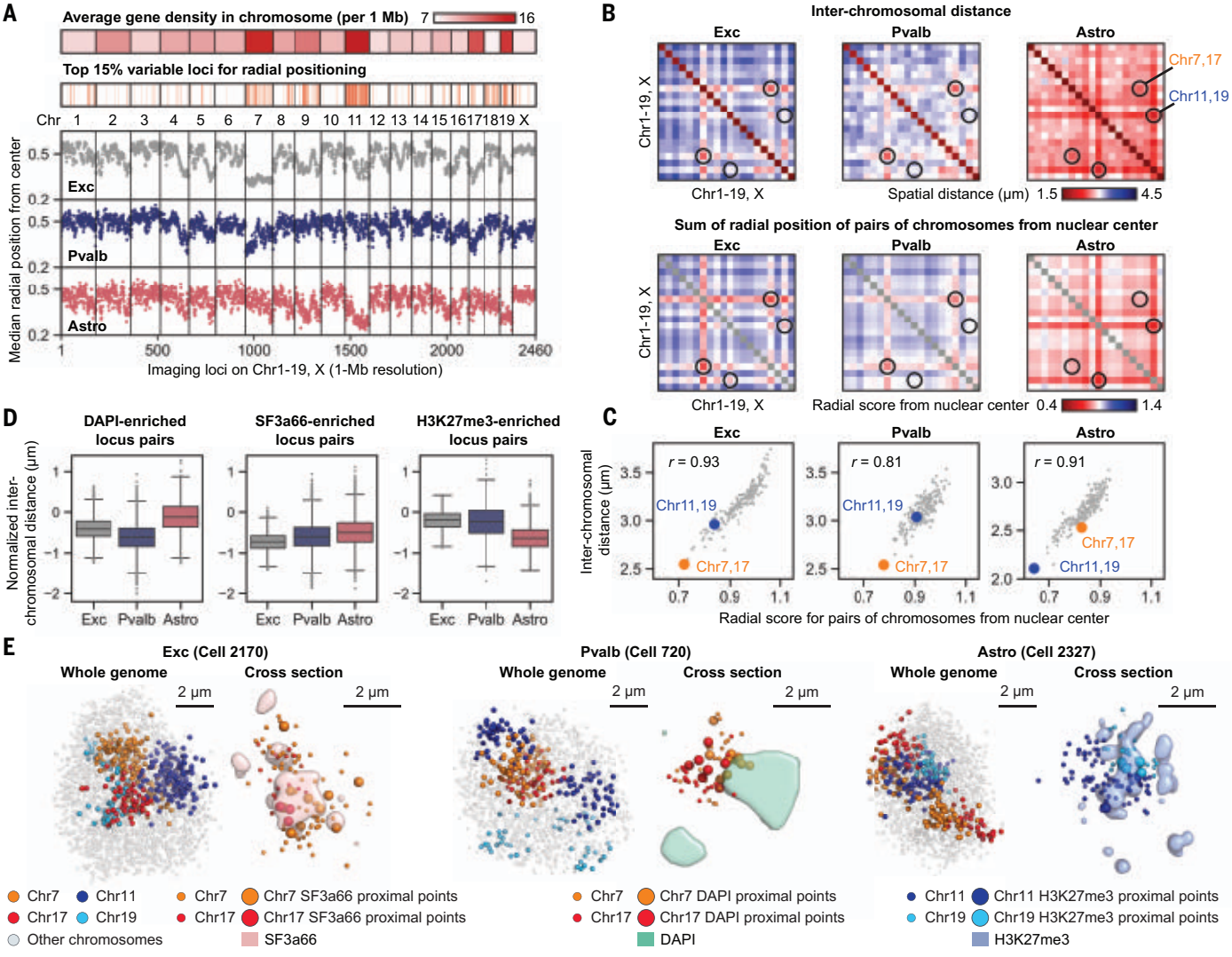


Fig. 5. Proximal points and nuclear bodies organize nuclear architecture. (**A**) Median radial positioning from the nuclear center for all 2460 loci imaged at 1-Mb resolution for three major cell types. Gene density (top panel) and the top 15% variable loci in terms of radial positioning among the three cell types (middle panel) are shown. (**B**) Comparison of mean pairwise distance between chromosomes (top panels) and the radial position scores for pairs of chromosomes (bottom panels) for the three major cell types. (**C**) Scatter plot of mean pairwise distance between chromosomes and the sum of radial position scores for pairs of chromosomes shown in (B). (**D**) Normalized interchromosomal distance between pairs of proximal points enriched in association with specific nuclear bodies and chromatin marks. The distances are normalized by subtracting an averaged interchromosomal spatial distance from all interchromosomal pairs of loci in each cell type. The boxplots represent the median, interquartile ranges, whiskers within 1.5 times the interquartile range, and outliers. (**E**) Chromosomes 7 and 17 are closer to the nuclear interior in neurons, shown with the proximal points on nuclear speckles (excitatory neuron, left) and heterochromatin (Pvalb inhibitory neuron, middle). Chromosomes 11 and 19 are closer to the nuclear interior in the astrocyte with the proximal points on H3K27me3 globules (right). The cells are shown with 2-μm cross sections for the excitatory and Pvalb inhibitory neuron and with 1-μm cross section for the astrocyte for visual clarity. n = 1895, 155, and 152 cells for excitatory neurons, Pvalb inhibitory neurons, and astrocytes, respectively, from three biological replicates.

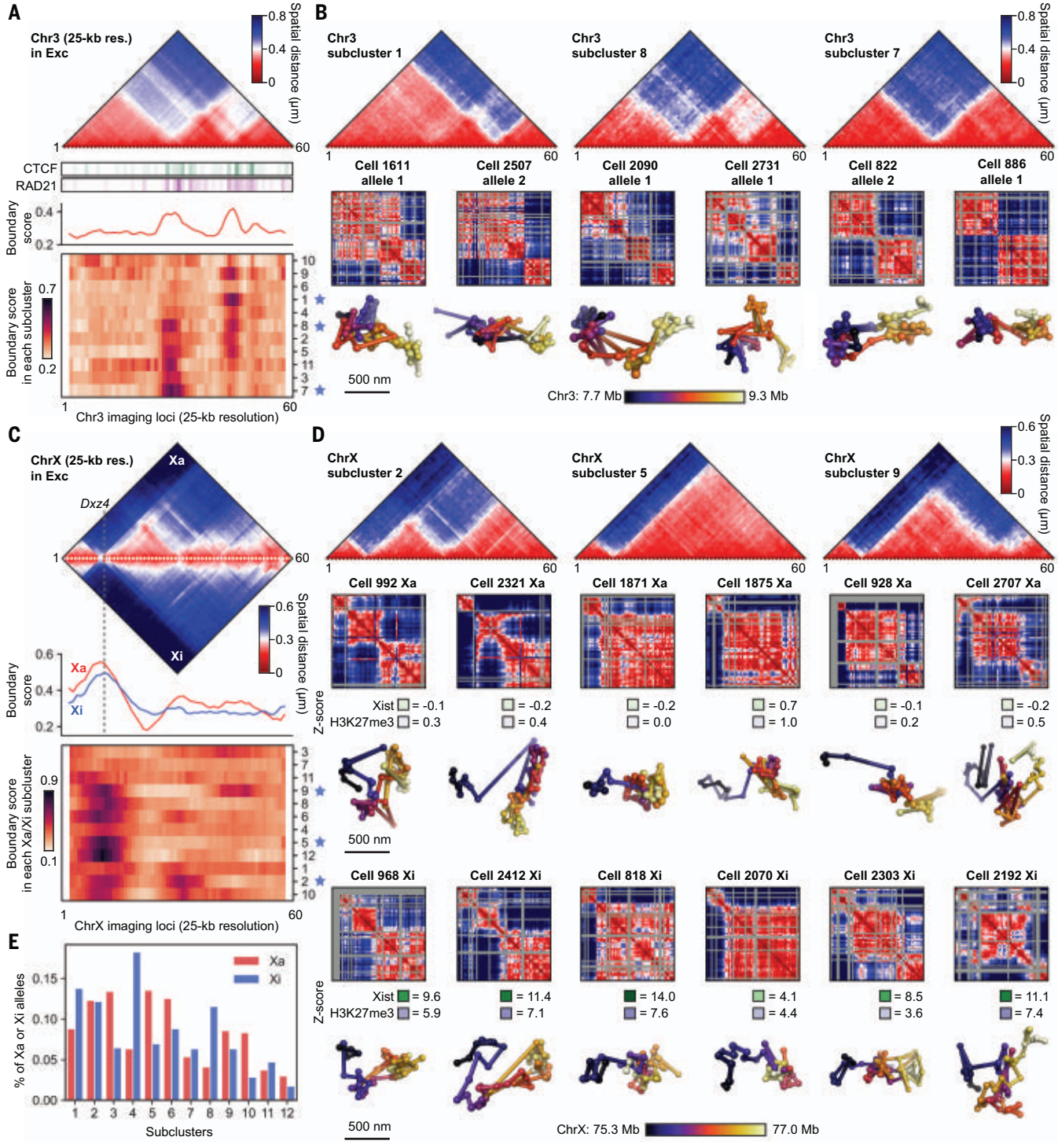


Fig. 6. Heterogeneous domain structures in single chromosomes. (**A**) Bulk averaged pairwise distance map from excitatory neurons (top) shows distinct domains in the chromosome 3 region, with boundaries and high median boundary score regions matching CTCF and cohesin binding sites. Single chromosomes were clustered into different domain structures with distinct median boundary score profiles (methods). (**B**) Examples for three of the structural subclusters are shown with different pairwise distance map structures. Single chromosome distance maps and chromosome structures from individual cells are shown below for each subcluster. Chromosomes with >20% loci detected in n = 1895 excitatory neurons from three biological replicates were used in (A) and (B). (**C**) Pairwise median distance maps for Xa (top, upper triangle) and Xi (top, lower triangle) in excitatory neurons from the 25-kb DNA seqFISH+ data at *Dxz4* locus (*33*). Clustering of the domain structures for individual Xa and Xi shown with the median boundary scores for Xa and Xi (middle) and for each subcluster containing Xa and Xi (bottom) in excitatory neurons. The subclusters highlighted with blue stars in (A) and (C) are shown in (B) and (D), respectively, as spatial distance maps. (**D**) Examples for subclusters with corresponding single-cell pairwise distance maps and single-cell 3D chromosome structures for Xa and Xi in excitatory neurons. Individual Xa and Xi with different underlying epigenetic states, shown with z-scored fluorescence intensity for Xist and H3K27me3 on the imaged chromosomal regions, can have similar domain structures. (**E**) The frequency of each subcluster occurring in single cells for Xa and Xi from 805 excitatory neurons. n = 805 cells with >20% loci detected for both Xa and Xi regions from three biological replicates in (C) to (E).

heterochromatic bodies tend to be in the nuclear interior in Pvalb inhibitory neurons (Fig. 5C), unlike the more distributed positioning in other cell types. Consequently, chromosome 17 tends to be radially positioned near the nuclear interior in Pvalb neurons (fig. S17, A to C). Similarly, because chromosomes 11 and 19 have many H3K27me3 proximal points in astrocytes (fig. S16E) and H3K27me3 globules tend to be in the interior of the astrocyte nuclei compared with neurons (fig. S17, A and C), chromosomes 11 and 19 tend to be observed near the interior and interact with other chromosomes in astrocytes, but less so in neurons (Fig. 5, A to C).

In addition, we observed a cell type dependence in the average interchromosome spatial distances (Fig. 5B, top, and fig. S17H). Notably, those pairwise interchromosome distance maps agree with averaged radial positioning of pairs of chromosomes, such that chromosomes in the nuclear interior tend to be spatially close to each other (Fig. 5, B, bottom, and C). Consequently, pairs of interchromosomal loci assigned as proximal points at nuclear bodies tend to be spatially closer to each other than the other nonproximal point pairs in neurons and astrocytes (Fig. 5D). Transcriptionally up-regulated loci between cell types tend to show closer interchromosomal distances and more interior radial positioning in the nucleus at the population level (fig. S18). However, in single cells, these loci do not show strong radial gradients (fig. S19), reflecting the variable nuclear speckle organization in individual nuclei and directly supporting a refined model of radial nuclear organization (*10*). These results show that the complexity in the global organization of chromosomes can be captured in the different nuclear body arrangements and proximal point chromatin profiles.

Domain boundaries are variable in single cells

Bulk-averaged chromosome conformation capture data revealed that chromosomes are organized into topologically associating domains (TADs) with clear insulation boundaries (*19*, *27*, *28*). More-recent imaging and sequencing experiments found TAD-like domain structures in single cells (*7*, *12*, *15*, *16*, *29*), whose boundaries are variable and shift stochastically across single cells upon depletion of cohesin (*12*).

To systematically investigate such single-cell chromosome domain structures across different chromosomal regions, we determined whether there are subpopulations of chromosomes with distinct configurations that differ from the bulk averaged configurations in excitatory neurons. We clustered the single chromosome pairwise spatial distance data with 25-kb resolution using principal components analysis (*30*) (fig. S20). Whereas the bulk data displayed three domains at the chromosome 3 region (7.7 to 9.3 Mb) (Fig. 6A, top), analysis of the single-cell data showed multiple configurations with different pairwise spatial associations and domain boundaries (Fig. 6, A, bottom, and B, and fig. S20). Chromosomes in single cells appeared to stochastically form domains preferentially from a subset of CTCF- and cohesin-binding sites (Fig. 6, A and B), consistent with a previous observation (*12*). We also observed single-cell domain structures at chromosomal regions without clear ensemble domain structures (*12*) (fig. S21), with characteristic ensemble high-order domain structures (*9*) (fig. S22), and with heterogeneous nucleolar associations (fig. S23), confirming the prevalence of single-cell domain structures.

Active and inactive X chromosome organization in single cells

We further examined the differences between the active X chromosome (Xa) and the inactive X chromosome (Xi) (*22*) in the female mouse brain cortex. The imaging-based genomics data can straightforwardly distinguish the Xa and Xi on the basis of their mutually exclusive associations with Xist RNA, a long noncoding RNA that is specifically expressed from and associates with the Xi (*22*), and characterize distinct epigenetic states between the Xa and Xi (fig. S24, A and B).

We observed that the Xa and Xi have distinct median distances between pairs of loci at different genomic length scales and in different cell types (fig. S24, C to H). Although the Xi is more compact than the Xa globally at the larger scale of tens of megabases (fig. S24, C and G) (*11*, *31*), we found, using the population-averaged spatial distance quantification, that the Xa appears to be more structured and compact at megabase or smaller length scales (fig. S24, C, D, G, and H).

Both the Xa and Xi have heterogeneous domain structures in individual cells (Fig. 6, C and D, and fig. S25). Even the region of the Xi that appears unstructured in the bulk data from both DNA seqFISH+ (Fig. 6C) and allele-specific Hi-C studies (*32*, *33*) appeared to adopt discrete domains in subsets of cells (Fig. 6D). We found that similar domain subclusters can be used in both the Xa and Xi, but with different relative frequencies for specific chromosome conformation (Fig. 6E), which leads to different average conformations for the Xa and Xi (Fig. 6C). A recent observation of heterogeneous single-cell domain structures in the Xa and Xi even under epigenetic perturbations (*34*) further supports our finding in the Xa and Xi. Taken together, although the Xa and Xi show very different epigenetic states and ensemble-averaged chromosome conformation, they can share similar underlying single-cell domain structures with different relative conformational preferences at the submegabase scale.

Discussion

The integrated spatial measurements of RNA, DNA, and chromatin marks in the same cell in the mouse brain enable the observation of proximal points, DNA loci that are present at the exterior of nuclear bodies and protein globules, across diverse cell types in tissues. Because each chromosome contains a distinctive pattern of proximal points associated with nuclear bodies and chromatin marks, these proximal points form a scaffold of the chromosomes on the exterior of the nuclear bodies in single cells. In addition, as the proximal points profiles and nuclear body organization are cell type dependent, these proximal points lead to distinct chromosomal positioning and interchromosomal relationships in the nucleus in each cell type. These findings extend our previous observations in mouse ES cells (*17*) and suggest a general principle of nuclear organization.

The integrated measurements also allow us to correlate the nuclear organization with transcriptional states in different cell types. We found that physical proximity to nuclear speckles and H3K27me3 are related to transcriptional changes in the tissues, extending recent work in cell cultures (*25*). In contrast, we observed that the submegabase domain structures in the Xa and Xi can be similar in single cells despite their large epigenetic and transcriptional differences, suggesting a potential decoupling between single-cell domain structures and underlying epigenetic states for the X chromosome.

The robust demonstration of integrated spatial genomics in tissues indicates that the same approach can be applied to a diverse range of biological systems to further explore the diversity and invariants in single-cell nuclear architecture. This integrated technology can be further scaled up using signal amplification methods (*35*–*37*) to allow larger tissue sections to be imaged. In addition, super-resolution imaging of epigenetic markers overlaid on the super-resolved DNA seqFISH+ data would provide finer resolution in single-cell chromatin profiles (*38*). We anticipate addressing many biological questions in future studies using genome-scale chromosome imaging together with transcriptome-scale profiling (*39*, *40*) and super-resolution protein imaging (*38*).

REFERENCES AND NOTES

1. T. Stuart, R. Satija, *Nat. Rev. Genet.* **20**, 257–272 (2019).
2. C. Zhu, S. Preissl, B. Ren, *Nat. Methods* **17**, 11–14 (2020).
3. T. Misteli, *Cell* **183**, 28–45 (2020).
4. J. Dekker *et al.*, *Nature* **549**, 219–226 (2017).
5. R. Kempfer, A. Pombo, *Nat. Rev. Genet.* **21**, 207–226 (2020).
6. A. M. Cardozo Gizzi *et al.*, *Mol. Cell* **74**, 212–222.e5 (2019).
7. L. J. Mateo *et al.*, *Nature* **568**, 49–54 (2019).

8. C. Zhu *et al.*, *Nat. Methods* **18**, 283–292 (2021).
9. S. A. Quinodoz *et al.*, *Cell* **174**, 744–757.e24 (2018).
10. Y. Chen *et al.*, *J. Cell Biol.* **217**, 4025–4048 (2018).
11. S. Wang *et al.*, *Science* **353**, 598–602 (2016).
12. B. Bintu *et al.*, *Science* **362**, eaau1783 (2018).
13. M. Liu *et al.*, *Nat. Commun.* **11**, 2907 (2020).
14. H. Q. Nguyen *et al.*, *Nat. Methods* **17**, 822–832 (2020).
15. J.-H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, *Cell* **182**, 1641–1659.e26 (2020).
16. A. C. Payne *et al.*, *Science* **371**, eaay3446 (2021).
17. Y. Takei *et al.*, *Nature* **590**, 344–350 (2021).
18. Y. Takei, N. Pierson, S. Shah, J. White, L. Cai, DNA seqFISH+ tissue analysis code and probe sequences, version 1, Zenodo (2021); https://doi.org/10.5281/zenodo.5507507.
19. J. R. Dixon *et al.*, *Nature* **485**, 376–380 (2012).
20. B. Tasic *et al.*, *Nat. Neurosci.* **19**, 335–346 (2016).
21. I. Solovei *et al.*, *Cell* **137**, 356–368 (2009).
22. R. Galupa, E. Heard, *Annu. Rev. Genet.* **52**, 535–566 (2018).
23. Y. Takei, L. Cai, Supplementary files for integrated spatial genomics in mouse brain cortex, version 1, Zenodo (2021); https://doi.org/10.5281/zenodo.4708112.
24. R. Imai *et al.*, *Mol. Biol. Cell* **28**, 3349–3359 (2017).
25. L. Zhang *et al.*, *Genome Res.* **31**, 251–264 (2020).
26. B. McStay, I. Grummt, *Annu. Rev. Cell Dev. Biol.* **24**, 131–157 (2008).
27. E. P. Nora *et al.*, *Nature* **485**, 381–385 (2012).
28. T. Sexton *et al.*, *Cell* **148**, 458–472 (2012).
29. M. V. Arrastia *et al.*, *Nat. Biotechnol.* (2021).
30. A. N. Sawh *et al.*, *Mol. Cell* **78**, 96–111.e6 (2020).
31. K. Teller *et al.*, *Nucleus* **2**, 465–477 (2011).
32. A. Minajigi *et al.*, *Science* **349**, aab2276 (2015).
33. L. Giorgetti *et al.*, *Nature* **535**, 575–579 (2016).
34. Y. Cheng, M. Liu, M. Hu, S. Wang, *bioRxiv* [Preprint] (2021). https://doi.org/10.1101/2021.05.12.443887.
35. S. Shah, E. Lubeck, W. Zhou, L. Cai, *Neuron* **92**, 342–357 (2016).
36. S. H. Rouhanifard *et al.*, *Nat. Biotechnol.* **37**, 84–89 (2019).
37. J. Y. Kishi *et al.*, *Nat. Methods* **16**, 533–544 (2019).
38. M. A. Woodworth *et al.*, *Nucleic Acids Res.* **49**, e82 (2021).
39. S. Shah *et al.*, *Cell* **174**, 363–376.e16 (2018).
40. C. L. Eng *et al.*, *Nature* **568**, 235–239 (2019).

ACKNOWLEDGMENTS

We thank N. Ollikainen and M. Guttman for help with Hi-C analysis; R. J. Oakey for providing the processed ChIP-seq data; C.-H. L. Eng for helpful discussion; C. Karp for making custom-made flow cells; L. Sanchez and E. Buitrago-Delgado for help with the perfusion setup; I. Strazhnik for help with figures; and A. Anderson for help with the manuscript. **Funding:** This work was supported by NIH 4DN DA047732 and supplement (L.C.) and by the Paul G. Allen Frontiers Foundation Discovery Center (L.C.). **Author contributions:** Y.T. and L.C. conceived of the idea and designed experiments. Y.T. designed probes with input from C.H.T. Y.T., J.Y., and S.Sc. prepared and validated experimental materials. Y.T. performed imaging experiments. Y.T., S.Sh., N.P., and J.W. wrote image processing scripts, and Y.T. performed image processing. Y.T., S.Z., and L.C. analyzed data with input from G.-C.Y. Y.T. and L.C. wrote the manuscript with input from S.Z., C.H.T., and G.-C.Y. L.C. supervised all aspects of the projects. **Competing interests:** L.C. is a cofounder of Spatial Genomics, Inc. **Data and materials availability:** The custom-written scripts and seqFISH probe sequences used in this study are available in a Zenodo repository (*18*) associated with a GitHub repository (https://github.com/CaiGroup/dna-seqfish-plus-tissue). The source data and processed data from this study are available in a Zenodo repository (*23*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj1966
Materials and Methods
Figs. S1 to S25
Tables S1 to S4
References (*41*–*81*)
MDAR Reproducibility Checklist

26 April 2021; accepted 15 September 2021
Published online 30 September 2021
10.1126/science.abj1966

REPORTS

PLANKTON

Global drivers of eukaryotic plankton biogeography in the sunlit ocean

Guilhem Sommeria-Klein[1,2]*, Romain Watteaux[3]†, Federico M. Ibarbalz[1]‡, Juan José Pierella Karlusich[1], Daniele Iudicone[3], Chris Bowler[1], Hélène Morlon[1]*

Eukaryotic plankton are a core component of marine ecosystems with exceptional taxonomic and ecological diversity, yet how their ecology interacts with the environment to drive global distribution patterns is poorly understood. In this work, we use *Tara* Oceans metabarcoding data, which cover all major ocean basins, combined with a probabilistic model of taxon co-occurrence to compare the biogeography of 70 major groups of eukaryotic plankton. We uncover two main axes of biogeographic variation. First, more-diverse groups display clearer biogeographic patterns. Second, large-bodied consumers are structured by oceanic basins, mostly through the main current systems, whereas small-bodied phototrophs are structured by latitude and follow local environmental conditions. Our study highlights notable differences in biogeographies across plankton groups and investigates their determinants at the global scale.

Marine plankton communities play key ecological roles at the base of oceanic food chains and in driving global biogeochemical fluxes (*1*, *2*). Understanding their spatial patterns of distribution is a long-standing challenge in marine ecology that has lately become a key part of the effort to model the response of oceans to environmental changes (*3*–*6*). Part of the difficulty lies in the constant recirculation of plankton communities by ocean currents (*7*), along which many physical, chemical, and biological processes—the so-called seascape (*8*)—modify community composition (*9*). Recent planetary-scale ocean sampling expeditions have revealed that eukaryotic plankton are taxonomically and ecologically extremely diverse, possibly even more so than prokaryotic plankton (*10*). Eukaryotic plankton cover an exceptional range of sizes, from picosized (0.2 to 2 μm) to mesosized (0.2 to 20 mm) organisms and larger, and ecological functions, including phototrophy (e.g., Bacillariophyta), phagotrophy (e.g., Ciliophora), mixotrophy (e.g., Dinophyceae), parasitism (e.g., marine alveolates), grazing (e.g., copepods), and filter feeding (e.g., tunicates). How these differences in body size and ecological function modulate the influence of oceanic currents and local environmental conditions on geographic distributions is poorly known (*11*–*13*).

In this research, we studied plankton biogeography across all major eukaryotic groups in the sunlit ocean by using 18S-V9 ribosomal DNA (rDNA) metabarcoding data from the *Tara* Oceans global survey [including recently released data from the Arctic Ocean (*14*)]. The data encompass 250,057 eukaryotic operational taxonomic units (OTUs) that were sampled at the surface and at the deep chlorophyll maximum (DCM) across 129 stations. We use a probabilistic model that allows identification of sets of OTUs that tend to co-occur across samples, which we coin "assemblages" (*15*, *16*); see materials and methods (*17*). A given OTU may belong to several assemblages, and each local community (i.e., the set of OTUs found in a given station) can be seen as a sample drawn in various proportions from the assemblages. Across the *Tara* Oceans samples and considering all eukaryotic OTUs together, we identified 16 geographically structured assemblages, each composed of OTUs that cover the full taxonomic range of eukaryotic plankton (Fig. 1, figs. S1 to S4, and supplementary text S1). Local planktonic communities often cannot be assigned to a single assemblage, as would be typical for terrestrial macro-organisms on a fixed landscape (*18*, *19*), but are instead mixtures of assemblages (Fig. 1A). This is consistent with previous findings that suggest that neighboring plankton communities are continuously mixed and dispersed by currents (*9*, *13*).

In addition to the global biogeography represented in Fig. 1, we investigated the biogeography of the 70 most-diversified deep-branching groups of eukaryotic plankton by fitting our probabilistic assemblage model separately to each group (Fig. 2; figs. S5 and S6, A to C; and

[1]Institut de Biologie de l'École Normale Supérieure (IBENS), École Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France. [2]Department of Computing, University of Turku, Yliopistonmäki, 20014 Turku, Finland. [3]Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy.
***Corresponding author. Email: guilhem.sk@gmail.com (G.S.-K.); helene.morlon@bio.ens.psl.eu (H.M.)**
†Present address: CEA, DAM–Île de France (DIF), 91297 Arpajon, France.
‡Present address: Universidad de Buenos Aires, CONICET, Centro de Investigaciones del Mar y la Atmósfera (CIMA), C1428EGA Buenos Aires, Argentina.

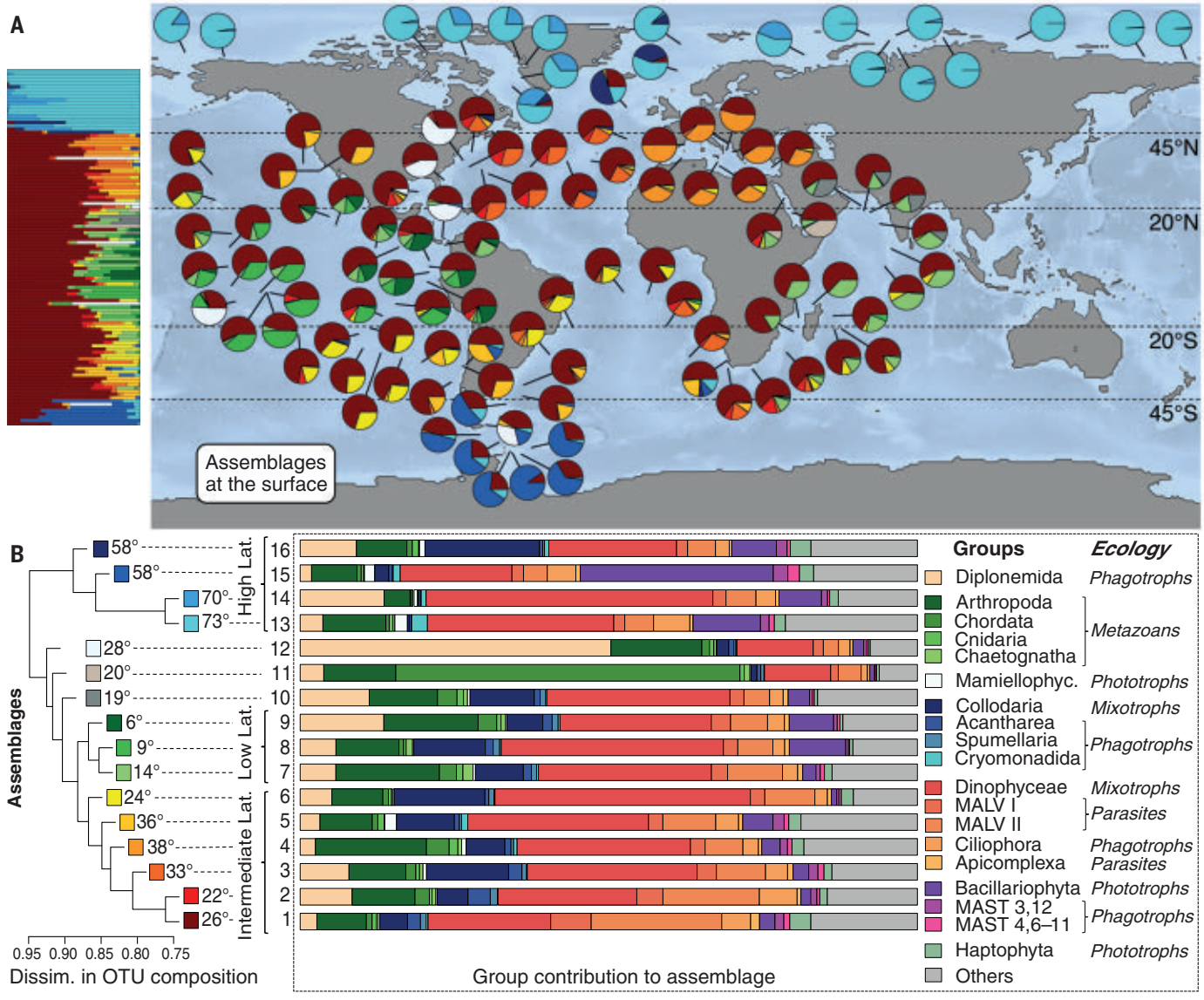


Fig. 1. Global surface biogeography of all eukaryotic plankton OTUs. (**A** and **B**) Biogeography is characterized by 16 assemblages of co-occurring OTUs that are identified by color [in (A) and left side of (B)] and a number from 1 to 16 [in (B)]. Shown in (A) are the relative contributions of assemblages to surface plankton communities, which are represented as pies on the map and as stacked bars ordered by latitude on the left-hand side. Most communities are dominated by one of three assemblages: parasite-rich assemblage 1 (dark red), which is globally ubiquitous except in the Arctic, and high-latitude, diatom-rich assemblages 13 and 15 (cyan and marine blue), with the transition occurring around the shift between cold and warm waters in the North and South Atlantic (subtropical fronts, at about 45°). Shown on the left in (B) is a dendrogram of assemblage dissimilarity with respect to their OTU composition (Simpson dissimilarity), where the mean absolute latitude of each assemblage is indicated. Assemblages form three main clusters that correspond to high, intermediate, and low latitudes (with approximate boundaries at 45° and 20°), although three assemblages (10, 11, and 12) cluster apart. Shown in the box on the right is the contribution of the 19 major eukaryotic groups tallying more than 1000 OTUs to assemblages, grouped by phylogenetic relatedness. The dominant ecological function is indicated for each group. Assemblages 11 and 12 capture outlier stations, which are, respectively, Chordata enriched (likely linked to tunicate blooms) and Diplonemida enriched (likely linked to deep water influxes). See supplementary text S1 for more detail. MALV, marine alveolates; MAST, marine stramenopiles.

table S1). We compared these biogeographic patterns on the basis of the similarity in assemblage proportions across stations (figs. S7 and S6D and materials and methods). Eukaryotic plankton groups vary along two main axes reflecting the strength and the nature of their biogeographic patterns (Fig. 2, fig. S8, and materials and methods). The group position on the first axis is tightly correlated to short-distance spatial autocorrelation [Pearson correlation coefficient (ρ_P) = 0.91 at the surface; fig. S9A] and anticorrelated to the Shannon entropy of assemblage proportions (fig. S9, C to F). Groups that score low on this axis are thus characterized by many coexisting assemblages and strong local variation, or "patchiness," whereas those that score high have clearer, more-structured biogeographic patterns, which are characterized by a small number of dominant assemblages and similar composition in nearby communities. The group position on the second axis is positively correlated to the scale of biogeographic organization, which we measured as the characteristic distance at which spatial autocorrelation vanishes (ρ_P = 0.54, $P = 10^{-6}$ at the surface) (fig. S10A) and which ranges from ~7000 to

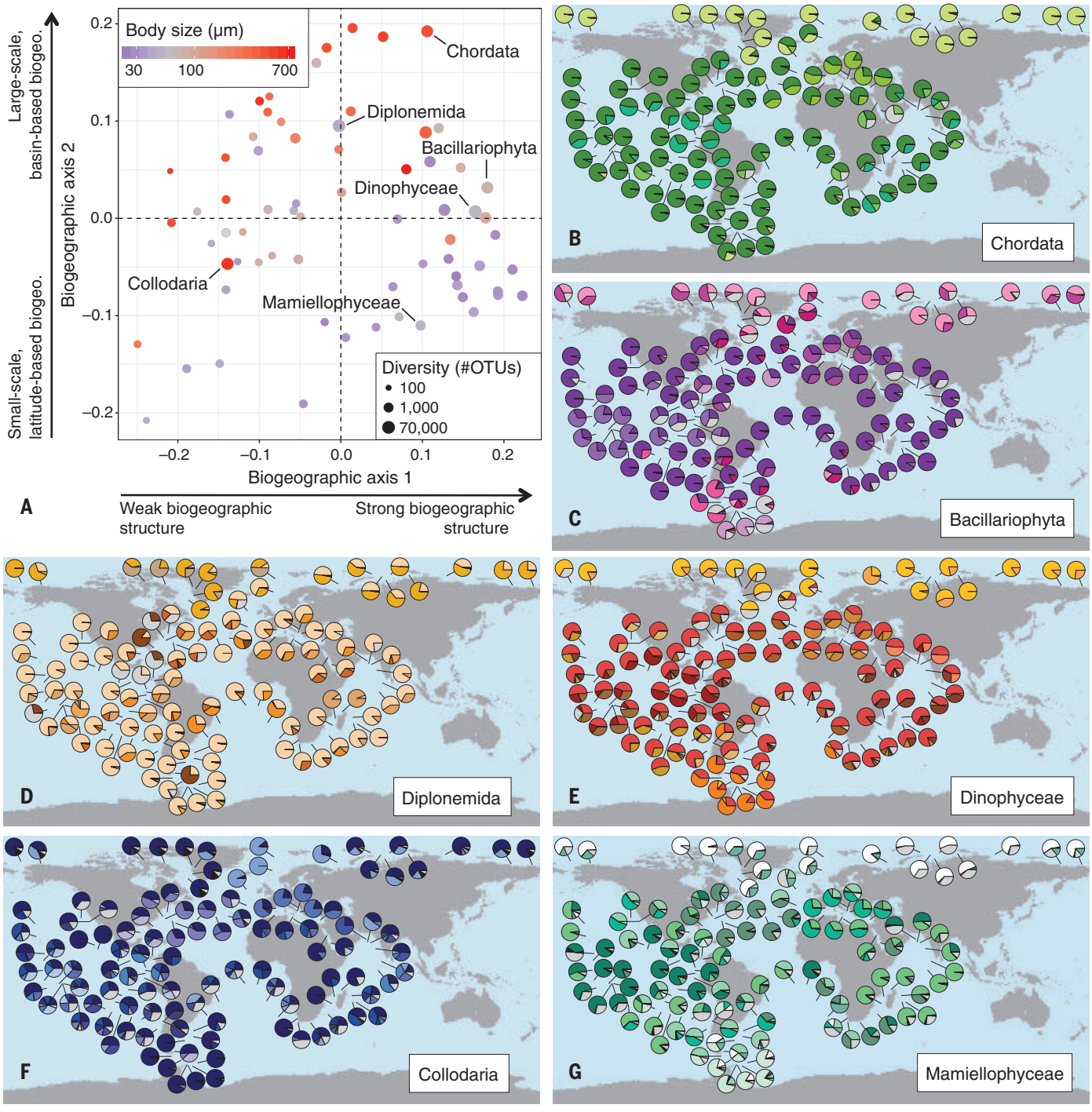


Fig. 2. Biogeographic heterogeneity across major eukaryotic plankton groups. (**A**) Principal coordinates analysis of the biogeographic dissimilarity between 70 major plankton groups. Each dot corresponds to the projection of a group onto the first two axes of variation; dot size varies with the group's log diversity, and dot color varies with its log mean body size. The 70 groups cover the full spectra of biogeographies, from those with weak spatial structure (left) to those organized at large spatial scale by oceanic basin (top right) or at smaller spatial scale and according to latitude (bottom right). (**B** to **G**) Surface biogeography of six representative groups that are labeled in (A). Only the most prevalent assemblages in each group are shown by using colors that are variations of the group's color in Fig. 1B; the remaining assemblages are left in gray. See fig. S5 for additional groups, and see (*31*) for assemblage-by-assemblage representations of the same biogeographies.

~14,400 km across groups at the surface. It is also positively correlated to within-basin autocorrelation ($\rho_P = 0.56$, $P = 10^{-7}$ at the surface; fig. S10C), which measures the tendency for communities from the same oceanic basin to be composed of the same assemblages (hereafter referred to as a "basin-based structure"), and negatively correlated with latitudinal autocorrelation ($\rho_P = -0.49$, $P = 10^{-5}$ at the surface; fig. S10E), which measures the tendency for communities at the same latitude on both sides of the equator to be composed of the same assemblages (hereafter referred to as a "latitude-based structure"; see materials and methods). Results are similar at the DCM, although less

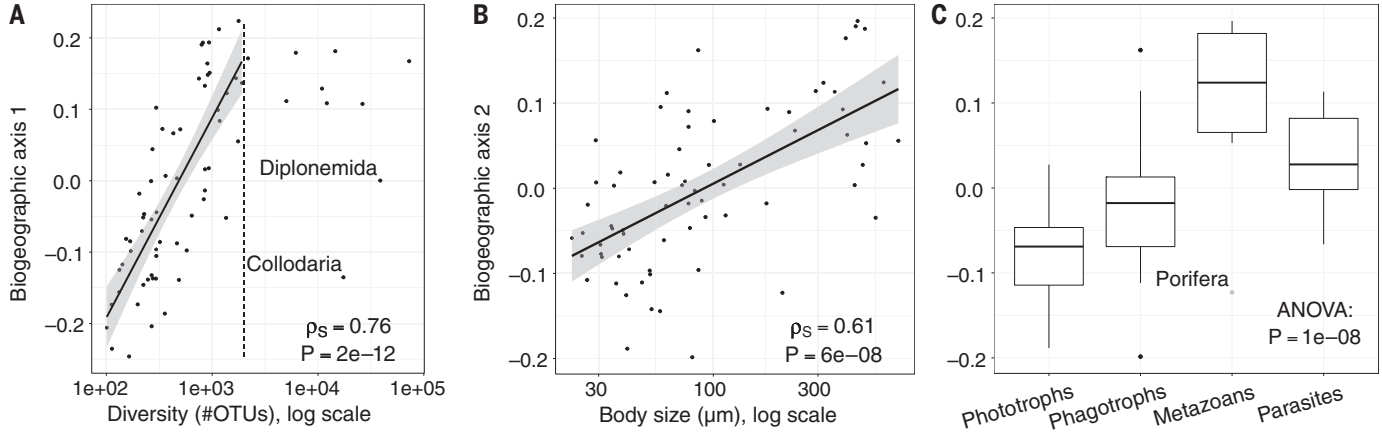


Fig. 3. Relationship between biogeography and diversity, mean body size, and ecological function across major eukaryotic plankton groups. (**A**) Group position along the first axis of biogeographic variation, which is indicative of the strength of biogeographic structure (see Fig. 2A) and increases sharply with diversity up to about 2000 OTUs (ρ_S, two-sided *t* test) but not beyond (as exemplified by Diplonemida and Collodaria, two of the most diverse groups). (**B**) Group position along the second axis increases with mean body size, which indicates that large-bodied plankton are organized at larger spatial scale and according to oceanic basins rather than latitude. (**C**) Positions along the second axis were binned into four broad ecological categories (*N* = 16, 28, 12, and 12, respectively; Collodaria and Dinophyceae were not categorized and are therefore not represented). Metazoan groups score high (except Porifera sponges, probably at the larval stage, which we excluded from statistical tests), phototrophs score low, and phagotrophs occupy an intermediate position (ANOVA, *F* test). Parasites score just below metazoans, which may reflect shared biogeography because of the importance of metazoans as hosts.

pronounced, possibly because of the smaller number of available samples and wider range of sampled depths (figs. S9 and S10). Although V9 OTUs may aggregate taxa with distinct biogeographies (*10*, *13*), this should weaken biogeographic differences across planktonic groups rather than generate spurious ones. In addition, we found consistent patterns with the alternative marker genes 18S-V4 rDNA and *psbO*, which suggests that our results are robust to variations in taxonomic resolution and coverage (see supplementary text S2). The notable differences that we observed across planktonic groups suggest that accounting for their specificities is crucial to understanding their biogeography (Fig. 2 and fig. S5).

We investigated how biogeographic differences among major groups relate to their diversity (number of OTUs), mean body size, and dominant ecological function, the latter of which we coarsely categorized as phototroph, phagotroph, metazoan, or parasite (table S1 and materials and methods). We found that the strength of biogeographic structure (group position on the first axis of variation; Fig. 2A) is strongly correlated to diversity [Spearman correlation coefficient (ρ_S) = 0.76 below 2000 OTUs; Fig. 3A]. This may be partly explained by the increased smoothing of stochastic variations and the larger amount of statistical information that is available in more-diverse groups, yet we found that the correlation between short-distance spatial autocorrelation and original group diversity persists after rarefaction to the same number of OTUs (fig. S11). This suggests that the generation and maintenance of eukaryotic plankton diversity are influenced by biogeographic structure, perhaps, in relation to endemism. Position on the first axis is also weakly anticorrelated to group mean body size ($\rho_S = -0.45$, $P = 10^{-4}$; fig. S12A) and, after controlling for diversity, is lower for metazoans than for phototrophs [analysis of covariance (ANCOVA) two-sided *t* test: $P = 0.04$; fig. S12D]. This agrees with the expected impact of turbulent stirring, which has been shown to generate patchier distributions in larger plankton (*20*, *21*). By contrast, the nature of biogeographic patterns (group position on the second axis; Fig. 2A) is strongly correlated to group mean body size ($\rho_S = 0.61$; Fig. 3B) and ecological function [analysis of variance (ANOVA) *F* test: $P = 10^{-8}$; Fig. 3C] and only weakly correlated to diversity ($\rho_S = 0.20$, $P = 0.1$); fig. S12E). Although mean body size covaries with ecological function (phagotrophs are larger than phototrophs on average, and metazoans are substantially larger than other plankton types; fig. S13), the positive relationship between group position on the second axis and mean body size still holds within ecological categories (ANCOVA *F* test: $P = 0.006$; fig. S14). Conversely, ecological function significantly influences group position on the second axis even after accounting for body size differences (ANCOVA *F* test: $P = 0.03$). These trends hold when directly considering measures of biogeographic structure (such as short-distance spatial autocorrelation and scale of biogeographic organization) rather than position on the two axes of variation (figs. S15 and S16) and when accounting for within-group variations in body size (fig. S17). In conclusion, diverse groups are more biogeographically structured, with large-bodied heterotrophs (metazoans such as copepods and tunicates) displaying a basin-based structure at the scale of oceanic basins or larger and small-bodied phototrophs (such as Mamiellophyceae and Haptophyta) displaying a latitude-based structure at a smaller spatial scale.

A global biogeography that matches oceanic basins suggests that communities are homogenized by ocean circulation at the basin scale (*13*) but have little ability to disperse between basins, either because of the comparatively limited connectivity by currents (*22*) or because of environmental barriers. Conversely, a biogeography that matches latitude and is symmetric with respect to the equator suggests a higher coupling to environmental variations (that are partly structured by latitude) and low between-basin dispersal limitation. To test these hypotheses, we used minimum transport times between pairs of stations, which were previously computed from a global ocean circulation model (*13*, *23*), and yearly averaged local environmental conditions. We transformed the matrix of minimum transport times into spatial patterns at different scales through eigenvector decomposition and thus obtained a set of so-called Moran eigenvector maps (hereafter simply referred to as "Moran maps"; see materials and methods). Each map represents the hypothetical geographic patterns expected for plankton with temporal variation along currents matching a specific spatial scale (figs. S18 and S19). We estimated local abiotic conditions by using yearly averaged values of temperature, nutrient concentration, and oxygen availability [World Ocean Atlas 2013 (*24*); see materials and methods]. Because biotic interactions (predation, competition, and parasitic and mutualistic symbiosis)

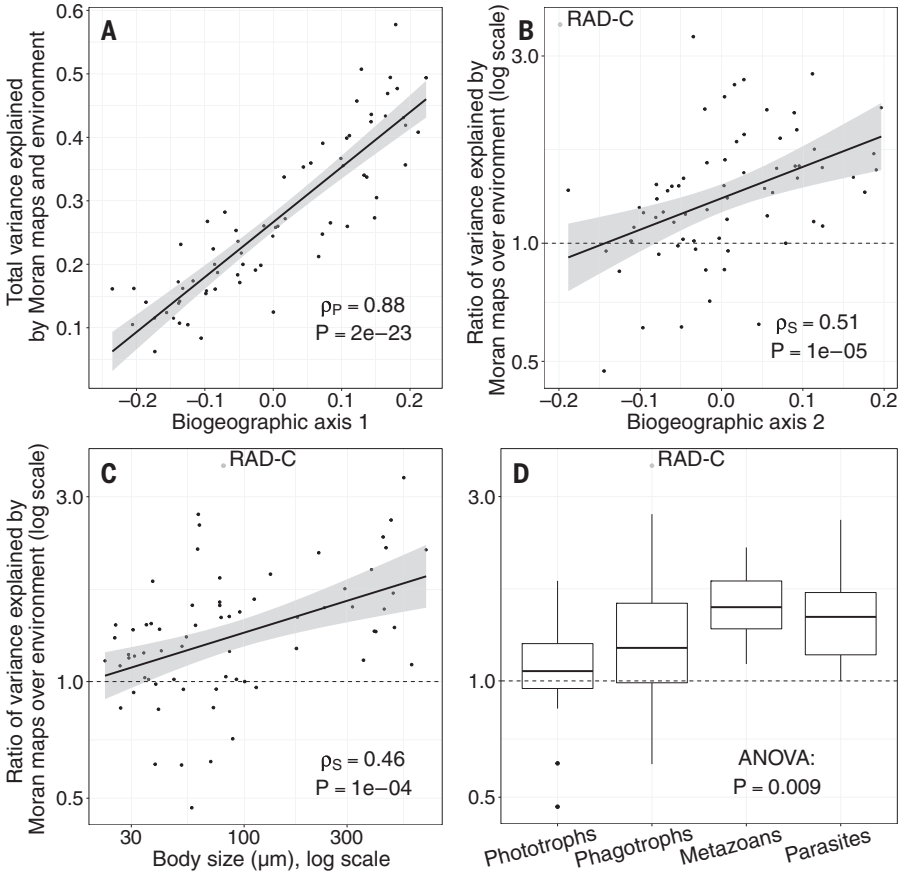


Fig. 4. Drivers of surface biogeography across major eukaryotic plankton groups. (**A**) The total variance in surface biogeography that can be explained by the combination of Moran maps and (abiotic and biotic) local environmental conditions tightly correlates with group position on the first axis of biogeographic variation (ρ_P, two-sided *t* test) (see Fig. 2A). (**B** to **D**) The ratio of the variance that is explained by Moran maps over the variance explained by local environmental conditions (B) increases with group position on the second axis (ρ_S, two-sided *t* test), (C) increases with group mean body size, and (D) varies across broad ecological categories (ANOVA, *F* test). The ratio is greater than 1 (dashed line) for most groups, which supports an overall stronger influence of processes occurring alongside transport than of local environmental conditions on plankton biogeography at the surface. The gray dot represents radiolarian group C (RAD-C), an outlier group that we excluded from statistical tests. We did not find any significant explanatory variable for Porifera and therefore excluded this group from these analyses.

are thought to be important determinants of plankton community composition (*25*), we also quantified local biotic conditions by using the relative read counts of major eukaryotic groups (excluding the focal group; see materials and methods). Biotic conditions, like abiotic ones, partly have a latitudinal structure, and we refer here to them collectively as "environmental conditions" (figs. S20 and S21). The resulting environmental maps can be interpreted as the hypothetical geographic patterns that are expected for organisms with a strong linear coupling to local environmental conditions. Therefore, the relative match of biogeography to environmental maps versus Moran maps indicates whether local environmental factors are able to linearly explain spatial variations in community composition or whether community composition results from a more complex combination of the physical and ecological processes that occur alongside transport: mixing, stirring, ecological drift (i.e., stochastic variations in abundances), and plankton interactions and responses to variations in temperature and nutrient supplies along currents.

We found that the total variance in surface community composition that can be statistically explained by Moran maps and local environmental conditions (abiotic and biotic) averages 27% across major eukaryotic groups (from 0% for Porifera to 58% for Ciliophora; see materials and methods). Expectedly, this total explained variance is tightly correlated to the strength of biogeographic structure (i.e., group position on first axis; $\rho_P = 0.88$; Fig. 4A). The variance that is statistically explained by Moran maps averages 21% across groups at the surface (fig. S22A) and is primarily contributed by maps corresponding to large-scale patterns between basins (figs. S19 and S23). This supports a key role of transport by the main current systems in shaping eukaryotic plankton biogeography, in particular by extending the distribution of taxa within basins (*9*, *13*, *26*, *27*) and constraining long-distance dispersal between basins (*22*). As expected from our previous results, the ratio of the fractions of variance that are statistically explained by Moran maps and local environmental conditions increases with group position on the second axis of variation ($\rho_S = 0.51$, $P = 10^{-5}$; Fig. 4B) and with group mean body size ($\rho_S = 0.46$, $P = 10^{-4}$; Fig. 4C) and depends on ecological function (ANOVA *F* test: $P = 0.009$; Fig. 4D). Local environmental factors can thus explain spatial variations in community composition better in phototrophs than in metazoans, the community composition of which appears to result from a more complex combination of the processes that occur alongside transport. Last, within the variance statistically explained by the local environment, an approximately equal share can be attributed to biotic and abiotic conditions for most groups (fig. S22B), irrespective of their mean body size, ecological function, diversity, or biogeography (fig. S24). The results are similar at the DCM but are far less pronounced (fig. S25). Although we cannot exclude the possibility that local biotic conditions reflect the indirect effect of local abiotic factors that are not accounted for in our study—such as fluxes of nutrients, which are often more relevant to planktonic organisms than instantaneous nutrient concentrations (*27*)—these results indicate an additional role for interspecific interactions in shaping community composition (*25*, *28*).

In the early days of the exploration of microeukaryote biogeography, the dominant view was that all eukaryotes up to a size of about 1 mm were globally dispersed and primarily constrained by abiotic conditions (*11*). That plankton of all sizes can be dispersal-limited is now well recognized, but the relationship between body size and biogeography remains unclear (*12*, *13*, *29*). It has been proposed that larger organisms tend to be more dispersal limited because of their lower population sizes and reduced ability to form stress-resistant propagules (*29*). However, compositional turnover along currents is slower, rather than faster, with increasing body size (*13*). Our results reconcile these views; at the global scale, organisms of larger sizes indeed have more-restricted ranges, whereas they have wider spatial distributions at the regional scale that are presumably linked to their specific ecologies, longer life spans, and reduced sensitivity to local environmental variations. Body size influences

biogeography through its close link with ecology and ultimately through the varying sensitivity of community size classes to abiotic and biotic environmental conditions as they drift along currents.

REFERENCES AND NOTES

1. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, *Science* **281**, 237–240 (1998).
2. A. Z. Worden *et al*., *Science* **347**, 1257594 (2015).
3. G. Beaugrand, R. R. Kirby, *Ann. Rev. Mar. Sci.* **10**, 169–197 (2018).
4. E. J. Raes *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8266–E8275 (2018).
5. D. Righetti, M. Vogt, N. Gruber, A. Psomas, N. E. Zimmermann, *Sci. Adv.* **5**, eaau6253 (2019).
6. D. P. Tittensor *et al*., *Nature* **466**, 1098–1101 (2010).
7. B. F. Jönsson, J. R. Watson, *Nat. Commun.* **7**, 11239 (2016).
8. M. T. Kavanaugh *et al*., *ICES J. Mar. Sci.* **73**, 1839–1850 (2016).
9. M. Lévy, O. Jahn, S. Dutkiewicz, M. J. Follows, *Limnol. Oceanogr.* **4**, 67–84 (2014).
10. C. de Vargas *et al*., *Science* **348**, 1261605 (2015).
11. B. J. Finlay, *Science* **296**, 1061–1063 (2002).
12. E. Villarino *et al*., *Nat. Commun.* **9**, 142 (2018).
13. D. J. Richter *et al*., bioRxiv 867739 [Preprint] (2020); https://doi.org/10.1101/867739.
14. F. M. Ibarbalz *et al*., *Cell* **179**, 1084–1097.e21 (2019).
15. G. Sommeria-Klein *et al*., *Mol. Ecol. Resour.* **20**, 371–386 (2020).
16. D. Valle, B. Baiser, C. W. Woodall, R. Chazdon, *Ecol. Lett.* **17**, 1591–1601 (2014).
17. Materials and methods are available as supplementary materials.
18. A. R. Wallace, *The Geographical Distribution of Animals: With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface, Volume 1* (Cambridge Univ. Press, 1876).
19. G. F. Ficetola, F. Mazel, W. Thuiller, *Nat. Ecol. Evol.* **1**, 89 (2017).
20. E. R. Abraham, *Nature* **391**, 577–580 (1998).
21. A. Bertrand *et al*., *Nat. Commun.* **5**, 5239 (2014).
22. F. L. Hellweger, E. van Sebille, N. D. Fredrick, *Science* **345**, 1346–1349 (2014).
23. S. Clayton *et al*., *Biogeosciences* **14**, 2877–2889 (2017).
24. T. P. Boyer *et al*., "World Ocean Database 2013," S. Levitus, A. Mishonov, Eds. (NOAA Atlas NESDIS 72, National Oceanographic Data Center, 2013); https://repository.library.noaa.gov/view/noaa/1291.
25. G. Lima-Mendez *et al*., *Science* **348**, 1262073 (2015).
26. M.-A. Madoui *et al*., *Mol. Ecol.* **26**, 4467–4482 (2017).
27. S. Dutkiewicz *et al*., *Biogeosciences* **17**, 609–634 (2020).
28. F. Vincent, C. Bowler, *mSystems* **5**, e00444-19 (2020).
29. J. B. H. Martiny *et al*., *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
30. *Tara* Oceans Consortium, Coordinators; *Tara* Oceans Expedition, Participants, Registry of all samples from the *Tara* Oceans Expedition (2009-2013), PANGAEA (2017); doi:10.1594/PANGAEA.875582.
31. G. Sommeria-Klein, guilhemSK/eDNA_LDA: Sommeria-Klein et al. Science 2021 release, Version 2.1, Zenodo (2021); doi:10.5281/zenodo.5557664.

ACKNOWLEDGMENTS

We thank O. Jaillon for feedback and help with *Tara* projects coordination; C. de Vargas for feedback and early discussions on the project; M. Follows and O. Jahn for sharing MITgcm simulation results for the Arctic Ocean; F. Hartig and O. Maliet for their guidance on Bayesian inference; L. Arístide and F. Delestro for their assistance with the figures; C. Fruciano and B. Perez for their advice on statistics; P. Frémont for his help with environmental data; B. B. Cael, L. Santoferrara, and an anonymous reviewer for helpful comments and constructive criticism; and F. Benedetti, J. Clavel, E. Kazamia, S. Lambert, E. Lewitus, M. Manceau, O. Missa, S. de Monte, I. Overcast, I. Quintero, E. Ser-Giacomi, A. C. Silva, and F. Vincent for suggestions and fruitful discussions. We also thank the commitment and support of the following people and sponsors: European Molecular Biology Laboratory (EMBL), Genoscope/CEA, agnès b., Etienne Bourgois, the Veolia Environment Foundation, Région Bretagne, Lorient Agglomération, World Courier, Illumina, the Électricité de France (EDF) Foundation, Fondation pour la recherche sur la biodiversité (FRB), the Prince Albert II de Monaco Foundation, and the *Tara* Ocean Foundation and its schooner and teams. **Funding:** This work was supported by European Research Council grants (ERC 616419-PANDA to H.M. and ERC 835067-DIATOMIC to C.B.), grants from the French Agence Nationale de la Recherche (MEMOLIFE, ref. ANR-10-LABX-54, to G.S.-K., H.M., and C.B. and OCEANOMICS, ref. ANR-11-BTBR-0008, to C.B.), funds from the MIUR Italian Flagship project RITMARE (to D.I.), a grant from the Sakari Alhopuro foundation (20210172 to G.S.-K.), and funds from CNRS. G.S.-K. received additional support from the Academy of Finland (decision 328791). C.B. and H.M. are members of the Research Federation for the study of Global Ocean Systems Ecology and Evolution (FR2022/*Tara* Oceans GOSEE). This article is contribution number 124 of *Tara* Oceans. **Author contributions:** G.S.-K. and H.M. designed the study with the help of R.W., D.I., and C.B. G.S.-K. performed the analyses. R.W. contributed the transport time data and their interpretation. F.M.I. assisted in the interpretation of 18S rDNA data. J.J.P.K. provided the *psbO* metagenomic data and their interpretation. G.S.-K. and H.M. wrote the paper with substantial input from all co-authors. **Competing interests:** The authors declare no competing interests. **Data availability:** All data reported here are available without restrictions. Metabarcoding data have been deposited at the European Nucleotide Archive under accession numbers PRJEB6610 and PRJEB9737 and metagenomes under project PRJEB402. Sample metadata are available from PANGAEA (*30*). All code, results and formatted input data are publicly available from Zenodo (*31*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abb3717
Materials and Methods
Supplementary Text
Figs. S1 to S32
Table S1
References (*32*–*55*)
MDAR Reproducibility Checklist

30 December 2020; accepted 23 September 2021
10.1126/science.abb3717

SEDIMENT TRANSPORT

Exceptional increases in fluvial sediment fluxes in a warmer and wetter High Mountain Asia

Dongfeng Li[1]*, Xixi Lu[1]*, Irina Overeem[2], Desmond E. Walling[3], Jaia Syvitski[2], Albert J. Kettner[2], Bodo Bookhagen[4], Yinjun Zhou[5], Ting Zhang[1]

Rivers originating in High Mountain Asia are crucial lifelines for one-third of the world's population. These fragile headwaters are now experiencing amplified climate change, glacier melt, and permafrost thaw. Observational data from 28 headwater basins demonstrate substantial increases in both annual runoff and annual sediment fluxes across the past six decades. The increases are accelerating from the mid-1990s in response to a warmer and wetter climate. The total sediment flux from High Mountain Asia is projected to more than double by 2050 under an extreme climate change scenario. These findings have far-reaching implications for the region's hydropower, food, and environmental security.

High Mountain Asia (HMA)—the Tibetan Plateau and the surrounding high Asian mountains—is Earth's third-largest ice reservoir and the origin of many of Asia's large rivers (*1*–*3*). Like the polar regions, HMA is experiencing amplified warming when compared with general global warming trends (*4*, *5*). The air temperature has warmed by ~2°C since the 1950s at a much faster rate than the global average (*4*). Precipitation across HMA has also increased over recent decades, although the increase is characterized by considerable spatial heterogeneity (*6*, *7*). Existing studies, based on large-scale cryospheric–hydrological models, have provided projections for future increases in the region's annual runoff due to climate change (*3*, *8*) and their potential impacts on agriculture, food, and ~2 billion people within HMA and downstream (*1*, *7*). One important potential consequence of these projected changes in runoff is the possibility of increased fluvial sediment fluxes across HMA, which would have important implications for hydropower generation and development, as well as water quality, and could affect the energy, food, and environmental security of the region (*2*, *9*, *10*). Here we investigate the impacts of the changing hydrology on fluvial sediment fluxes in this region.

We collate and analyze available flow and sediment load data for rivers in HMA, extending over the past six decades, to investigate changes in runoff and sediment flux in response to a warmer and wetter climate. To exclude the potential impact of human activities, we select 28 quasi-pristine headwater river basins with observations of both runoff and sediment flux (plus 14 basins that only have runoff observations) (table S1). The selection allows for the separation of climate change impacts on sediment flux from anthropogenic impacts (*2*, *11*, *12*). The dataset has not yet been included in current global discharge and sediment datasets (*2*, *9*, *10*, *13*). We examine the sensitivity of sediment flux to changing temperature and precipitation in HMA on the basis of observational data and a climate elasticity model (materials and methods). Projections for future sediment fluxes from HMA are offered with a qualitative assessment of downstream impacts.

[1]Department of Geography, National University of Singapore, Kent Ridge 117570, Singapore. [2]CSDMS, Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, CO 80309, USA. [3]Department of Geography, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, UK. [4]Institute of Geosciences, Universität Potsdam, 14476 Potsdam, Germany. [5]Changjiang River Scientific Research Institute, Wuhan 430010, China.
***Corresponding author. Email: dongfeng@u.nus.edu (D.L.); geoluxx@nus.edu.sg (X.L)**

Time series of observed runoff from the headwaters of HMA show substantial increases over the past six decades (5.06 ± 0.51%/10 years, mean ± SE) (Fig. 1 and table S2). The magnitude of these increases exhibits large spatial heterogeneities across the region, as well as between different stations within the same river basin. In the HMA-east basins (stations S1 to S11), the decadal trends of increasing runoff vary from 0.95 ± 1.19%/10 years at Shigu (Yangtze) to 12.08 ± 3.05%/10 years at Tuotuohe (Yangtze). In the HMA-west basins (stations S12 to S28), runoff exhibits significant decadal upward trends in most headwaters, with local rates of increase varying from 1.69 ± 1.39%/10 years at Yingxiongqiao (Tienshan North) to 8.94 ± 2.97%/10 years at Keerguti (Tarim).

The observed sediment fluxes from the headwaters of HMA exhibit larger rates of increase (12.99 ± 1.18%/10 years, mean ± SE) than runoff and again reveal distinct spatial heterogeneities (Fig. 1 and table S3). In the HMA-east basins, increases in sediment flux vary from 3.67 ± 5.71%/10 years at Jimai (Yellow) to 25.40 ± 4.92%/10 years at Daojiebe (Salween), whereas the equivalent rates of increase for runoff are only 2.56 ± 2.76%/10 years and 2.13 ± 1.30%/10 years, respectively. In the HMA-west basins, increases in sediment flux vary from 5.30 ± 2.04%/10 years at Shaliguilanke (Tarim) to 24.68 ± 8.57%/10 years at Yuzimenleke (Tarim), whereas the equivalent decadal rates of runoff increase are only 4.00 ± 1.52%/10 years and 9.07 ± 1.67%/10 years, respectively. The increase of water and sediment discharge accelerated after 1995 (Fig. 2, D and E, and fig. S7), which is consistent with the faster warming and wetting trend in HMA after the mid-1990s (Fig. 2, A and B).

Increasing sediment discharge appears to be related to the rapidly retreating glaciers, which can intensify glacial sediment production (*14*–*16*), although their spatial extent is relatively small (Figs. 2 and 3). Accelerating permafrost thaw appears to increase sediment flux (Fig. 2) as permafrost disturbance–related sediment sources [e.g., thaw slumps (fig. S1)] become active in a warming climate. The previously frozen landscapes then become prone to erosion and increased water and sediment fluxes (*17*). The large spatial heterogeneity associated with the magnitude of the increases in sediment flux can be attributed to variations

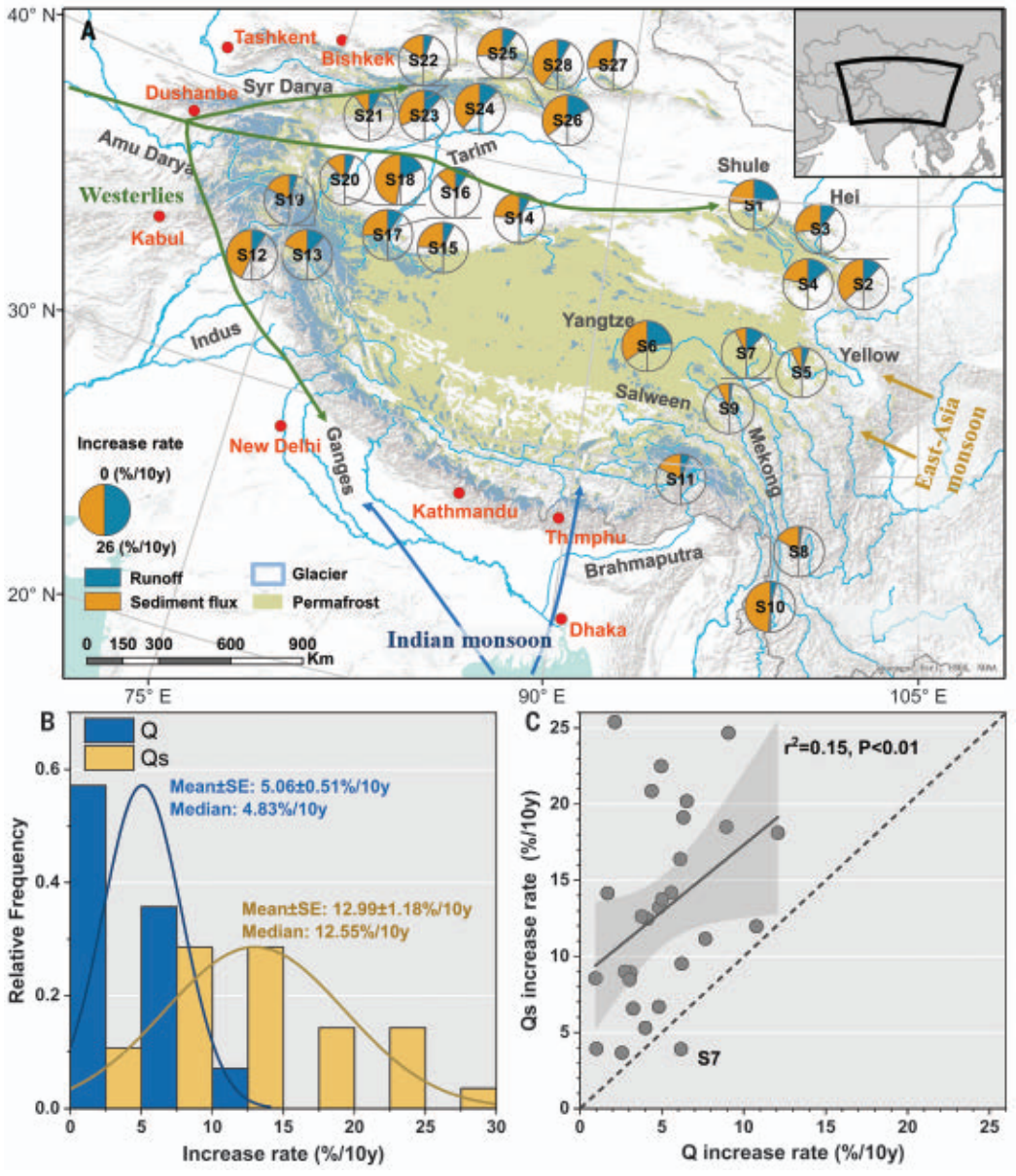


Fig. 1. Increasing annual runoff and fluvial sediment fluxes in a warming and wetting HMA. (**A**) Relative rates of increase of mean annual runoff and sediment flux over the past six decades. Stations S1 to S11 represent HMA-east basins (monsoon-dominated basins), and stations S12 to S28 represent HMA-west basins (westerly-dominated basins). S1 to S11 are also influenced by westerlies in some parts of their upstream basins (*33*, *34*). Details of the 28 headwater basins are in table S1. (**B**) Statistics on the rates of increase of annual runoff (Q) and sediment flux (Qs) (tables S2 and S3). (**C**) Relationship between the rate of increase of sediment flux and the rate of increase of runoff. Information on the distribution of glaciers and permafrost presented in (A) was sourced from (*34*, *35*). The dotted line represents the 1:1 line. The gray-shaded area denotes the 95% confidence interval of the best-fit line.

in the magnitude of changes in runoff (Fig. 1C) and is also likely to reflect the varying storage capacity of the sediment transfer pathways. In the upper Brahmaputra River (above S11), ~40% of the upstream annual sediment supply has been deposited in the wide valleys (*18*). Similarly, substantial amounts of sediment are deposited in the Kosi River floodplain, a tributary of the upper Ganges (*19*). Our findings provide robust evidence of landscape change in a warming climate similar to those reported for other cold environments (e.g., polar regions) (*15*, *20*).

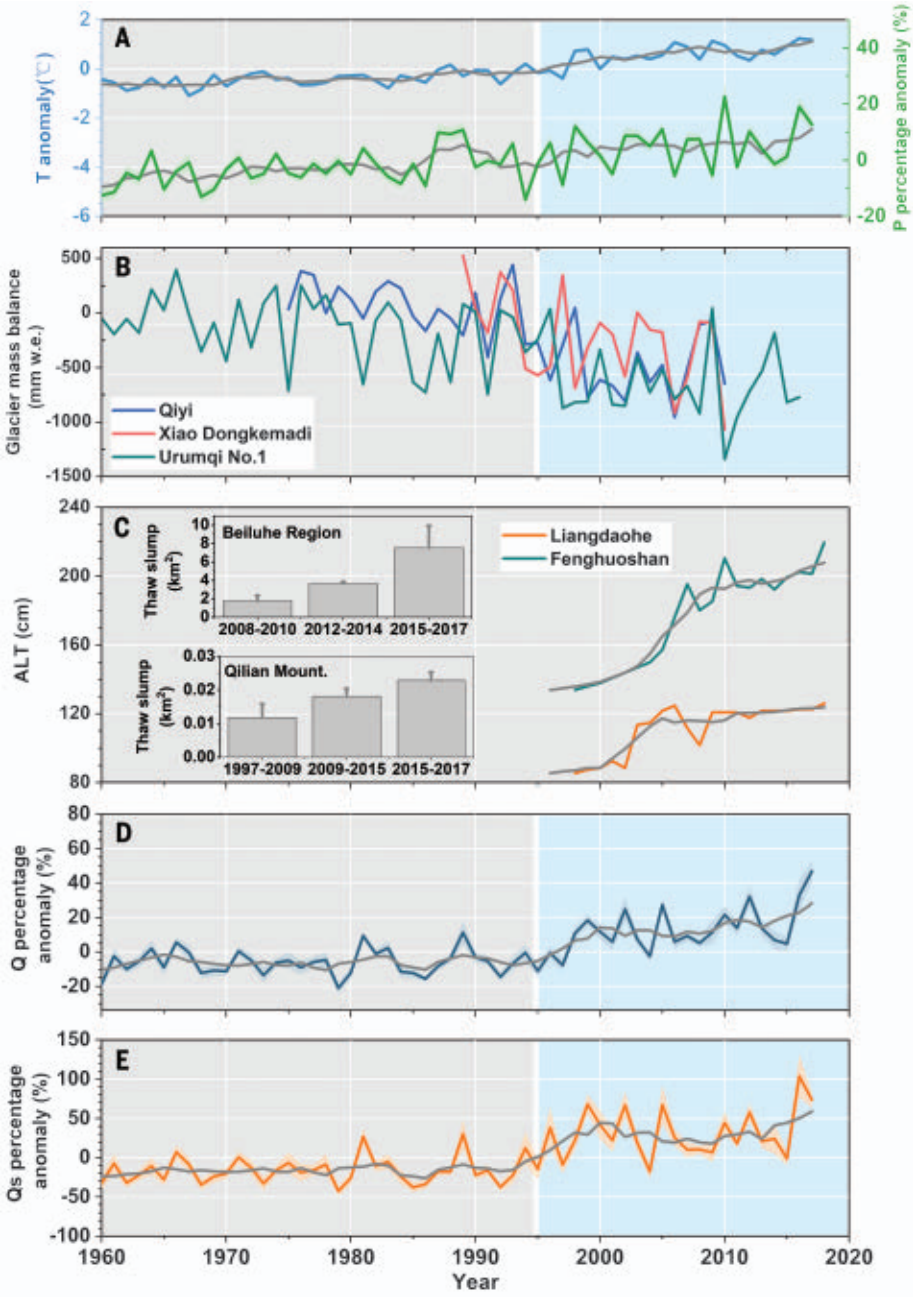


Fig. 2. Observed increases in the annual runoff and sediment flux in response to changes in air temperature, precipitation, glacier melt, and permafrost thaw. Gray lines denote the 5-year moving average for the data. (**A**) Increase in the annual mean air temperature (T) (0.32° ± 0.03°C/10 years, $P < 0.01$) and precipitation (P) (2.09 ± 0.53%/10 years, $P < 0.01$) anomalies over HMA based on 108 climate stations (materials and methods and fig. S12). (**B**) Increase in the glacier mass melt rates from Qiyi Glacier (near S3), Xiaodongkemadi Glacier (near S6), and Urumqi No. 1 Glacier (near S27), which represent three of the longest observed time series of glacier mass balance in HMA (*34*). w.e., water equivalent. (**C**) Increasing active-layer thickness (ALT) at Liangdaohe (near S6) and Fenghuoshan (near S6) (*35*) and the associated increase in thaw slumps (important sediment sources for permafrost environments) in the Beiluhe region and the Qilian Mountain (near S3 to S7) (*36*, *37*). (**D**) Increasing annual runoff (Q) anomaly (in percentage) for the entire HMA (mean ± SE of the 28 stations). (**E**) Increasing annual sediment flux (Qs) anomaly (in percentage) for the entire HMA (mean ± SE for the 28 stations). The light blue– and orange-shaded areas denote SEs.

Increasing fluvial sediment fluxes in HMA are mainly the result of accelerated thermally driven glacier-permafrost melt and increased precipitation and associated erosion processes (Fig. 2 and figs. S1 to S3) (*20*, *21*). To estimate the geomorphic impacts of climate change on HMA, we analyze the sensitivity of sediment fluxes to increases in temperature and precipitation, using a climate elasticity model based on the past six-decadal observations of climate, runoff, and sediment flux and a conceptual framework of runoff-sediment generation in cold environments (materials and methods and fig. S3). On average, a 10% increase in precipitation results in a 24 ± 5% increase in sediment flux, and a 1°C increase in air temperature results in a 32 ± 10% increase in sediment flux (Fig. 3 and materials and methods). Our data also show that the sensitivity of sediment flux to increasing precipitation decreases with increased glacierization (percentage of glacier cover for a headwater basin). By contrast, the sensitivity of sediment flux to increasing temperature generally increases with increased glacierization. The four outliers from this trend are possibly biased by low sediment connectivity and low erodibility (fig. S11). The sensitivity results provided by the elasticity model are in agreement with historical observations (fig. S10) and evidence a greater susceptibility to change than, for example, pan-Arctic rivers (*20*). The greater sensitivity of sediment flux to climate change in HMA as compared with that in the Arctic environments reflects some basic geomorphic facts: HMA has a high gradient of much softer lithologies as compared with lower-gradient Arctic environments and their hard craton lithologies (fig. S6) (*20*).

We provide projections for the increase of sediment flux from HMA by the middle of the 21st century in response to climate change, on the basis of the results of the sensitivity analysis and different future scenarios of climate change (Fig. 4 and table S4). By using an area-weighted–average approach and specific sediment-yield observations from 122 stations (materials and methods and table S5), the present-day fluvial sediment flux from HMA is estimated to be 1.94 ± 0.80 metric gigatons (Gt) per year [the area-weighted mean specific sediment yield is 410 ± 170 metric tons per square kilometer per year, and the area of HMA is 4.72×10^6 km^2 (*22*)]. For an extreme climate change scenario [i.e., temperature increases by 3°C and precipitation increases by 30% by 2050, relative to the period 1995–2015 (*4*, *23*)], the total sediment flux from the entire HMA will increase from the present 1.94 ± 0.80 Gt/year to 5.18 ± 1.64 Gt/year (Fig. 4 and table S4). However, the

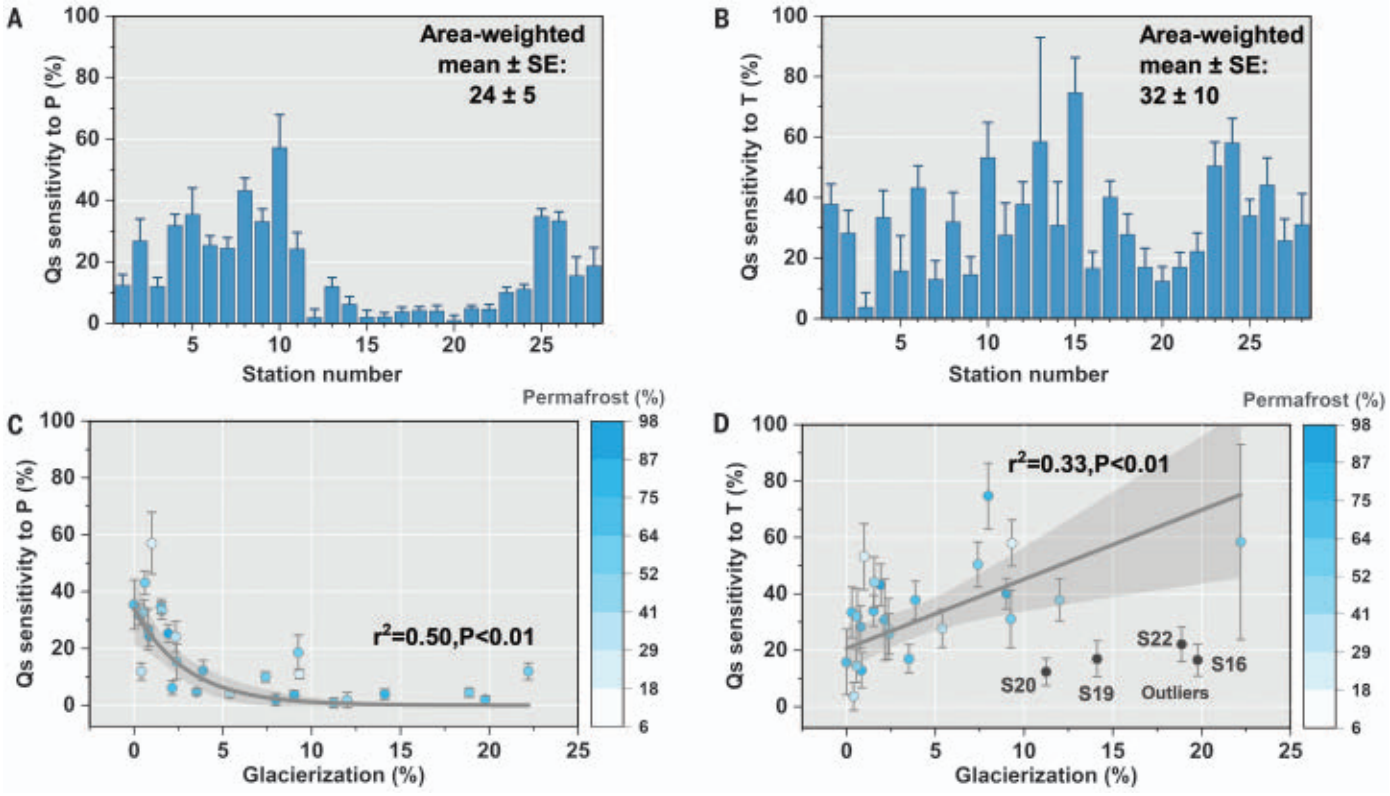


Fig. 3. Sensitivity of sediment flux to increased precipitation and temperature. (**A**) Sensitivity of sediment flux to precipitation (percentage change in sediment flux for every 10% increase in precipitation). (**B**) Sensitivity of sediment flux to air temperature (percentage change in sediment flux for every 1°C increase in air temperature). Each bar denotes one headwater basin ranked (from left to right) by the station number as shown in Fig. 1. (**C**) Sensitivity of sediment flux to precipitation significantly decreases with glacierization (percentage of glacier cover for a headwater basin). (**D**) Sensitivity of sediment flux to temperature shows a positive relationship with glacierization. The four outliers possibly reflect low sediment connectivity and low erodibility of their basins (fig. S11). The colored key represents the permafrost cover fraction for a basin. The permafrost cover fraction shows no significant correlations (r^2) with precipitation sensitivity ($r^2 < 0.1$, $P > 0.5$) or temperature sensitivity ($r^2 < 0.1$, $P > 0.5$), possibly because of the more important influence of the largely unknown permafrost ground ice conditions.

predicted future fluvial sediment flux from HMA represents projections that reflect the impact of climate change on sediment export from basin outlets located both within and toward the margins of HMA. As such, they are therefore not directly indicative of the actual sediment output from HMA, because sediment storage in downstream river reaches, future dam construction, and land-use changes could also affect the future sediment fluxes (*2*, *24*, *25*).

The projected major increase in fluvial sediment flux within HMA has important implications for hydropower projects, water quality, and food and energy security in both HMA and its downstream river systems. A recent study reported that damming the deglacierizing basins of HMA could contribute to the region's energy production and help it adapt to climate change by temporarily storing the glacier meltwater and modifying the flow regime (*26*). Hydropower potential in Nepal and Bhutan exceeds their present electricity consumption (*26*). However, increasing fluvial sediment fluxes could severely threaten existing and planned reservoirs in HMA and its downstream margins (fig. S12) by decreasing their storage capacities faster than anticipated and abrading power turbines. Increasing sediment flux from the headwaters of the Yangtze River over recent decades is reported to be affecting the cascade of reservoirs in the upper Yangtze River, China's largest hydropower production region (*21*, *27*). Under the extreme climate change scenario, we predict that from 2020 to 2050, a total of ~2000 metric megatons (Mt) of sediment will likely be deposited in existing and planned reservoirs located on the upper Yangtze River, equivalent to a reduction of storage capacity by ~1.5×10^9 m^3 (approximately two times the designed storage capacity of the downstream Liyuan Reservoir) (fig. S13). For the upper Indus River, the water storage capacity of the Tarbela Reservoir, the largest hydropower and irrigation project in Pakistan, decreased by >3.5×10^9 m^3 (~30% loss of original storage capacity and therefore much faster than the original expectation) between 1974 and 2006 (*28*), which threatens the expected life span of the reservoir, water supply, irrigation capacity, hydropower generation, flood control, and thereby food, energy, and environment security. Furthermore, in many rivers of HMA, sediment mostly comprises hard minerals such as quartz and feldspar that are known for causing abrasion and damaging turbine components. In the Sutlej River of the upper Indus basin, the turbines of the Nathpa Jhakri hydropower project (1500 MW) were replaced after <1 year of operation because of the severe abrasion problem (*28*).

Increasing runoff and sediment flux from HMA and associated increases in nutrient, pollutant, and carbon fluxes will affect not only water availability and storage but also water quality and flooding in the densely populated downstream locations (fig. S3). Previous studies have reported increasing levels of mercury released from melting glaciers and thawing permafrost in recent years across the Tibetan Plateau (*4*). Additionally, fine suspended

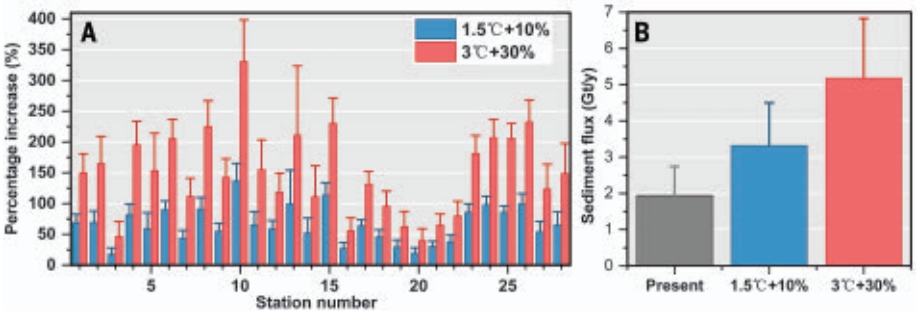


Fig. 4. Projection of future sediment fluxes in HMA. (**A**) Percentage increases in the sediment flux at each station under the conservative scenario (1.5°C + 10%; temperature increases by 1.5°C and precipitation increases by 10% by the middle of the 21st century relative to the period of 1995–2015) and the extreme scenario (3°C + 30%; temperature increases by 3°C and precipitation increases by 30% by the middle of the 21st century relative to the period 1995–2015). Each bar denotes one headwater basin ranked (from left to right) by the station number as shown in Fig. 1. (**B**) Projected increases in the sediment flux from the entire HMA from the present 1.94 ± 0.80 Gt/year to 3.32 ± 1.18 Gt/year (conservative scenario) and to 5.18 ± 1.64 Gt/year (extreme scenario) by the middle of the 21st century. Projections of other climate change scenarios are in table S4.

sediment is an important vector for the transport of phosphorus and most heavy metals (e.g., chromium, arsenic, and lead) because of its high specific surface area and associated reactivity (*2*, *10*). Thus, increasing fluvial sediment flux from HMA is likely to increase sediment-associated nutrient and contaminant fluxes, which can negatively affect water quality and aquatic ecosystems. Furthermore, suspended sediment is a key vector for organic carbon transport (*29*); and the precise role of erosion and sediment delivery in mobilizing organic carbon from melting permafrost landscapes and delivering it to the fluvial system remains uncertain and represents an important research need. Quantifying this process is important when assessing the positive feedback between climate warming, permafrost degradation, and carbon cycling (*30*). Assuming a sediment carbon content of 1 to 3% (*9*), the total carbon flux associated with the sediment output from HMA could reach 50 to 150 Mt/year by 2050 under the extreme climate change scenario. A substantial proportion of the increased sediment flux could be temporarily deposited in the river system (e.g., wide alluvial valleys and river floodplains) (*18*) and aggrade riverbeds (*19*), potentially triggering river avulsions (*31*) and increasing the risks of flooding, particularly during the summer melt season. Nevertheless, the increasing fluvial sediment flux can also provide more aggregates for construction, electronic materials, and other related services (*32*).

This study provides observational evidence of increasing water and sediment fluxes in the headwaters of HMA and the associated spatial variation of the rates of increase between individual basins and between tributaries within basins. Increasing sediment fluxes will negatively affect existing and planned hydropower projects, influence irrigation capacity, and therefore threaten the region's food and energy security. The increasing fluvial sediment flux and its associated nutrients, pollutants, and organic carbon also have implications for water quality and flooding, potentially affecting millions of people in HMA and downstream regions. This study sheds light on the importance and potential implications of the marked increases in recent and future sediment fluxes that have not been fully recognized by scientific communities, such as the Intergovernmental Panel on Climate Change (IPCC), and the region's policy-makers, nor have they been fully taken into account in the assessment of potential changes in the global carbon cycle.

REFERENCES AND NOTES

1. W. W. Immerzeel *et al.*, *Nature* **577**, 364–369 (2020).
2. J. Best, *Nat. Geosci.* **12**, 7–21 (2019).
3. W. W. Immerzeel, L. P. H. van Beek, M. F. P. Bierkens, *Science* **328**, 1382–1385 (2010).
4. Intergovernmental Panel on Climate Change, *Special Report on the Ocean and Cryosphere in a Changing Climate* (Intergovernmental Panel on Climate Change, 2019).
5. B. J. Peterson *et al.*, *Science* **298**, 2171–2173 (2002).
6. T. Yao *et al.*, *Bull. Am. Meteorol. Soc.* **100**, 423–444 (2019).
7. P. Wester, A. Mishra, A. Mukherji, A. B. Shrestha, *The Hindu Kush Himalaya Assessment* (Springer, 2019).
8. M. Huss, R. Hock, *Nat. Clim. Chang.* **8**, 135–140 (2018).
9. J. P. M. Syvitski, C. J. Vörösmarty, A. J. Kettner, P. Green, *Science* **308**, 376–380 (2005).
10. D. E. Walling, *Geomorphology* **79**, 192–216 (2006).
11. M. Meybeck, L. Laroche, H. H. Dürr, J. P. M. Syvitski, *Global Planet. Change* **39**, 65–93 (2003).
12. J. P. M. Syvitski, A. Kettner, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **369**, 957–975 (2011).
13. L. Li *et al.*, *Sci. Bull. (Beijing)* **65**, 62–69 (2019).
14. F. Herman, F. De Doncker, I. Delaney, G. Prasicek, M. Koppes, *Nat. Rev. Earth Environ.* **2**, 422–435 (2021).
15. I. Overeem *et al.*, *Nat. Geosci.* **10**, 859–863 (2017).
16. A. A. Beylich, K. Laute, J. E. A. Storms, *Geomorphology* **287**, 126–143 (2017).
17. D. Li, I. Overeem, A. J. Kettner, Y. Zhou, X. Lu, *Water Resour. Res.* **57**, e2020WR028193 (2021).
18. X. Shi *et al.*, *Earth Surf. Process. Landf.* **43**, 432–443 (2018).
19. R. Sinha *et al.*, *J. Hydrol. (Amst.)* **570**, 156–166 (2019).
20. J. P. M. Syvitski, *Polar Res.* **21**, 323–330 (2002).
21. D. Li, Z. Li, Y. Zhou, X. Lu, *Geophys. Res. Lett.* **47**, e2020GL087745 (2020).
22. W. T. Pfeffer *et al.*, *J. Glaciol.* **60**, 537–552 (2014).
23. P. D. A. Kraaijenbrink, M. F. P. Bierkens, A. F. Lutz, W. W. Immerzeel, *Nature* **549**, 257–260 (2017).
24. H. Gupta, S.-J. J. Kao, M. Dai, *J. Hydrol. (Amst.)* **464–465**, 447–458 (2012).
25. O. Varis, M. Kummu, C. Lehr, D. Shen, *Appl. Geogr.* **53**, 105–116 (2014).
26. D. Farinotti, V. Round, M. Huss, L. Compagno, H. Zekollari, *Nature* **575**, 341–344 (2019).
27. D. Li, X. X. Lu, X. Yang, L. Chen, L. Lin, *Geomorphology* **322**, 41–52 (2018).
28. G. W. Annandale, G. L. Morris, P. Karki, *Extending the Life of Reservoirs: Sustainable Sediment Management for Dams and Run-of-River Hydropower* (The World Bank, 2016).
29. L. Ran *et al.*, *Nat. Commun.* **12**, 1730 (2021).
30. T. Wang *et al.*, *Sci. Adv.* **6**, eaaz3513 (2020).
31. J. P. M. Syvitski, G. R. Brakenridge, *GSA Today* **23**, 4–10 (2013).
32. M. Bendixen, J. Best, C. Hackney, L. L. Iversen, *Nature* **571**, 29–31 (2019).
33. A. Bhattacharya *et al.*, *Nat. Commun.* **12**, 4133 (2021).
34. T. Yao *et al.*, *Nat. Clim. Chang.* **2**, 663–667 (2012).
35. L. Zhao *et al.*, *Permafr. Periglac. Process.* **31**, 396–405 (2020).
36. C. Mu *et al.*, *Landslides* **17**, 1051–1062 (2020).
37. J. Luo, F. Niu, Z. Lin, M. Liu, G. Yin, *Geomorphology* **341**, 79–85 (2019).

ACKNOWLEDGMENTS

We thank J. Best and Z. Li for earlier comments and field assistance, respectively. **Funding:** This study was supported by the MOE (grants R-109-000-273-112 and R-109-000-227-115). D.L. acknowledges IPCC Scholarship Award (jointly founded by the IPCC and Cuomo Foundation) and President's Graduate Fellowship. I.O. acknowledges NSF OPP awards (2001225 and 1553172). **Author contributions:** D.L. and X.L. designed the study; D.L. performed all analyses with assistance from T.Z.; D.L. prepared the draft. All authors contributed to the manuscript and figures. **Competing interests:** The authors have no competing financial conflicts of interest with this study. **Data and materials availability:** The data reported in the paper are available at https://github.com/geolidf/HMA-catchments. Other data are presented in the supplementary materials. Annual runoff and sediment data are sourced from the hydrological data yearbooks published by the Ministry of Water Resources, China (http://www.mwr.gov.cn/english/), and the Water and Power Development Authority, Pakistan (http://www.wapda.gov.pk/). Climate data are sourced from China Meteorological Administration (http://data.cma.cn/en) and ERA-5 reanalysis product (https://cds.climate.copernicus.eu/). Data on glacier and permafrost are available at the Randolph Glacier Inventory (RGI 6.0; https://www.glims.org/RGI/) and the National Tibetan Plateau Data Center (http://www.tpdc.ac.cn/en/).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi9649
Materials and Methods
Figs. S1 to S13
Tables S1 to S5
Google Earth kmz file of catchment properties
References (*38–71*)

12 April 2021; accepted 16 September 2021
10.1126/science.abi9649

POLYMER CHEMISTRY

Polyethylene materials with in-chain ketones from nonalternating catalytic copolymerization

Maximilian Baur, Fei Lin, Tobias O. Morgen, Lukas Odenwald, Stefan Mecking*

The world's most abundantly manufactured plastic, polyethylene, consists of inert hydrocarbon chains. The introduction of reactive polar groups in these chains could help overcome problematic environmental persistence and enhance compatibility with other materials. We show that phosphinophenolate-coordinated nickel complexes can catalyze nonalternating copolymerization of ethylene with carbon monoxide to incorporate a low density of individual in-chain keto groups in polyethylene chains with high molecular weight while retaining desirable material properties. After processing by conventional injection molding techniques, tensile properties remain on par with those of standard high-density polyethylene while also imparting photodegradability.

Polyethylene is the most abundantly manufactured synthetic polymer, combining facile processing and low-cost production with beneficial mechanical properties (*1*). The latter arises from crystalline ordering of stretched hydrocarbon chains. This ordering is particularly pronounced for high-density polyethylene (HDPE), which consists of linear chains devoid of branching that would otherwise disturb crystalline packing (*2*). Polyethylene is hydrophobic and nonpolar and does not easily adhere to polar materials such as metal surfaces or wood. Because of the chemically inert nature of the hydrocarbon chains, polyethylene is not susceptible to degradation reactions and therefore persists when released to the environment.

One prospective approach to overcoming the drawbacks of crystalline polyethylene materials is catalytic copolymerization of ethylene with a low ratio of carbon monoxide. The resulting keto groups in the polyethylene chain could provide an array of desirable modes of reactivity—including photodegradability—as demonstrated for branched low-density polyethylenes (LDPEs) with ~1 mol % of keto groups from free-radical high-pressure copolymerization (*3*–*5*). Even in linear polyethylene, a low concentration of keto groups is unlikely to disturb the crystalline order (*6*).

However, such materials have remained elusive, despite numerous efforts toward copolymerization. Compared with copolymerization, alternative post-polymerization oxidation of polyethylene (*7*, *8*) requires an additional synthetic step and is not selective for keto functional groups; it also requires problematic reagents. In copolymerization, carbon monoxide exhibits much stronger binding to the catalyst compared with the ethylene monomer, which prevents consecutive incorporation of olefin (*9*). Thus, rather than a mildly perturbed polyethylene chain, alternating polyketones are formed, which are used in high-melting engineering thermoplastics with entirely different applications and processing properties than polyethylenes (*10*).

Here, we report that Ni(II) catalysts can achieve the long-sought nonalternating catalytic copolymerization of ethylene and CO. The resultant materials with desirably low content of keto groups in high–molecular weight polyethylene chains can be melt-processed and feature tensile properties similar to those of commercial HDPE.

For catalyst choice and design, traditional electron-deficient d^0-metal sites—such as those used in production of polyolefins (*1*)—are excluded as they are deactivated by polar molecules such as carbon monoxide. Quenching with carbon monoxide is an established method to deactivate such olefin polymerization catalysts (*11*). Catalytic processes involving CO, such as olefin carbonylation, generally rely on d^8-metal centers (*12*), and the aforementioned alternating polyketones are produced commercially with cationic palladium catalysts (*10*, *13*). Further, the strong binding affinity of CO—widely used as a ligand in organometallic chemistry—must be controlled. Irreversible displacement of other essential ligands, which destroys active sites, must be prevented. In addition, the coordination of CO or keto groups in the growing chain can potentially block coordination sites required for further chain growth. Notably, the kinetic preference for CO incorporation, which promotes the formation of alternating polyketones, needs to be overcome. In this sense, the target CO copolymerizations are the opposite of catalytic copolymerizations of ethylene with polar vinyl comonomers such as acrylates (*14*–*17*), in which the comonomer reactivity can be problematically low.

These arguments prompt consideration of neutral late transition metal catalysts with strongly bound chelating ligands as promising candidates. Compared with more common cationic polymerization catalysts, neutral active sites exhibit a less pronounced preference for binding of CO versus other ligands, including monomeric olefins. Indeed, the only precedent for consecutive ethylene insertions in CO copolymerization has been observed with neutral phosphinosulfonato palladium catalysts (*18*–*21*). However, low–molecular weight brittle waxes with number average

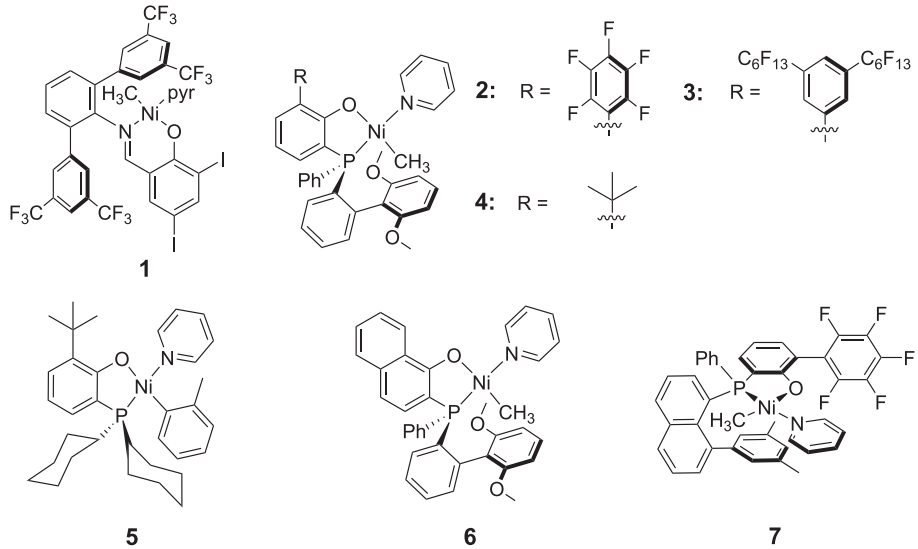


Fig. 1. Catalyst precursors studied. Salicylaldiminato complex **1**, and phosphinophenolato complexes **2** to **7**.

University of Konstanz, Department of Chemistry, Universitätsstraße 10, 78457 Konstanz, Germany.
*Corresponding author. Email: stefan.mecking@uni-konstanz.de

molecular weights $M_n \leq 3000$ g mol^{-1} (and likely heterogeneous compositions) were formed (*22*) and no polyethylene plastics could be obtained despite extensive studies by different joint industrial and academic ventures.

Neutral N,O-coordinated salicylaldiminato and P,O-coordinated phosphinophenolato Ni(II) catalysts in ethylene homopolymerization can produce linear semicrystalline polyethylene with a molecular weight of several million grams per mole (*23*). High polymerization temperatures are also expected to be favorable for nonalternating chain growth. For the general case of two competing pathways, the preference for the kinetically favored one decreases with increasing temperature according to $k_a/k_b \sim \exp(-\Delta\Delta G^{\neq}/RT)$ assuming a given difference of the energy barriers ($\Delta\Delta G^{\neq}$) for both pathways. R is the universal gas constant and T is temperature. The robust nature of P,O-coordinated nickel(II) catalysts is advantageous in this regard. They are applied at 140°C for industrial ethylene oligomerization (*24*). However, as a result of their propensity for chain transfer by β-hydride elimination events, traditional P,O-coordinated catalysts yield oligomers or low molecular weight polymers with M_n <10^4 g mol^{-1}, even at lower reaction temperatures (*25*–*27*). Only recently has this long-prevalent perspective been revised by the finding that appropriately bulky substituents can strongly suppress chain transfer to afford high molecular weight polyethylenes (*28*, *29*).

Notwithstanding the literature consensus that such catalysts are rapidly deactivated by CO and at most form only alternating polyketones with low activity (*26*, *30*), we exposed state-of-the-art salicylaldiminato (**1**) and phosphinophenolato (**2**) (Fig. 1) catalyst precursors to ethylene (E)/CO mixtures with low partial pressures of CO (Table 1) at polymerization temperatures found to be optimum in the established ethylene polymerizations with these catalysts. With (**1**), only small amounts of polymer were obtained even at an E/CO ratio of >10^2 (table S3, entries 11 and 12). Infrared (IR) spectra of the polymer were indicative of an alternating polyketone. By contrast, with (**2**), substantial amounts [up to 4 grams versus ~60 mg with (**1**)] of a polymer were obtained with a carbonyl band at a distinctly different wavelength than that of alternating polyketones (Table 1, entries 1 to 5, and Fig. 2B).

Encouraged by these findings, we explored a series of custom-modified, state-of-the-art neutral P,O-chelated catalyst precursors with bulky substituents on the phenolate moiety as well as aryl groups shielding one or both apical positions (Fig. 1; for synthesis details and full characterization data, see supplementary materials). Catalyst precursors (**2**) through (**6**) were all active in the presence of CO and, at high E/CO ratios on the order of 100, formed polymers that were nonalternating polyketones on the basis of IR spectral characterization (Table 1 and table S3).

The microstructures of the polymers were elucidated by ^{13}C nuclear magnetic resonance (NMR) spectroscopic methods. To enhance sensitivity, copolymers from pressure reactor experiments with ^{13}CO were used. By a combination of 1D and 2D NMR methods and reference to literature data, ^{1}H and ^{13}C NMR spectra could be assigned fully (figs. S16 to S18). This analysis revealed the presence of isolated ketone groups in the polyethylene chain (Fig. 2A). Further, nonalternating motifs of ketone groups in proximate positions to each other along the chain could be differentiated, and alternating motifs were observable. At desirable values of ~1% CO incorporation, most keto groups were nonalternating, and isolated keto groups were prevalent. The overall

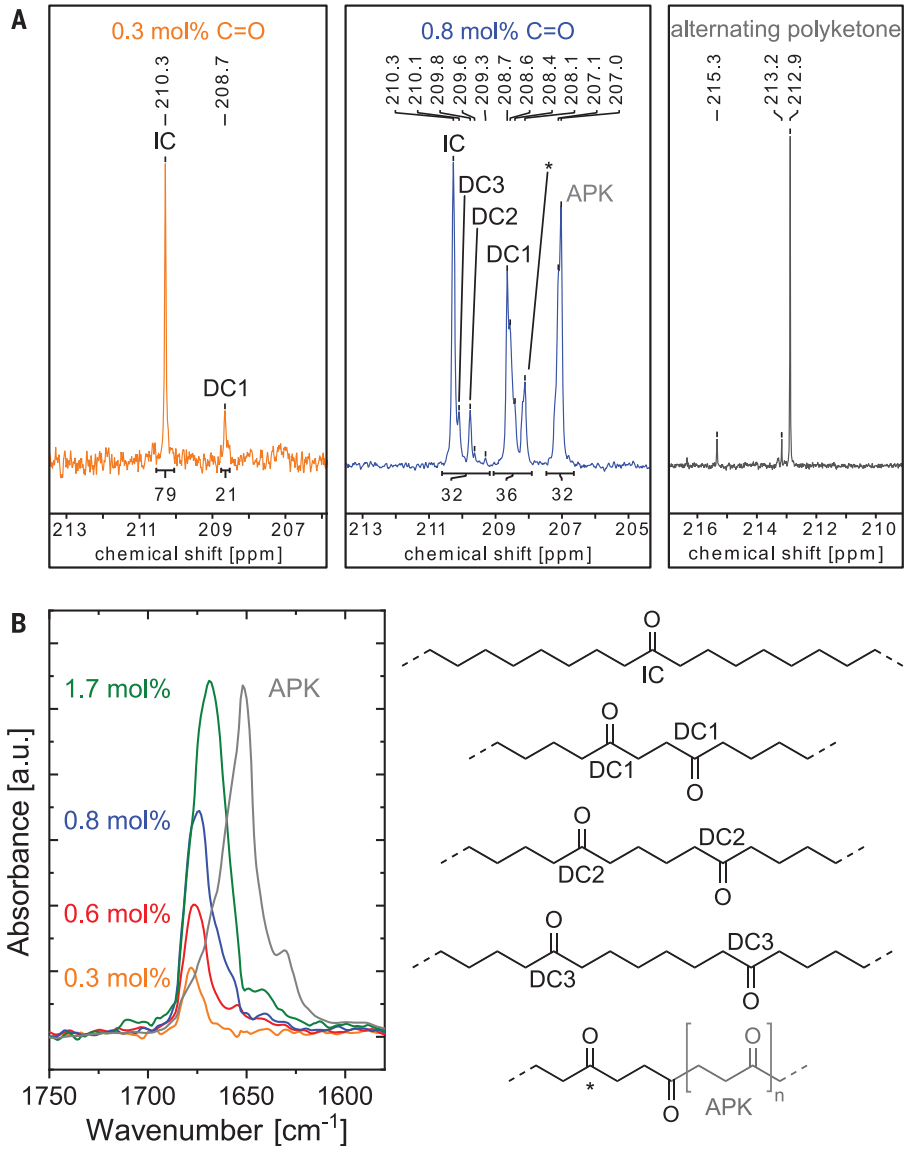


Fig. 2. Microstructure characterization of copolymers. (**A**) Carbonyl regime of ^{13}C NMR spectra of ^{13}CO copolymers with different keto contents ($C_2D_2Cl_4$ solvent, 383 K) and an alternating polyketone (hexafluoroisopropanol/C_6D_6 solvent, 300 K) for comparison. IC, Isolated carbonyl; DCX, Double carbonyl separated by X repeat units of ethylene; APK, Alternating polyketone; *, Carbonyl adjacent to an alternating sequence. (**B**) IR spectra of copolymers with different keto content, and alternating polyketone for comparison. The absorbance intensity of alternating polyketones is adjusted for clarity. Note these are ^{13}CO copolymers as in (A), which shifts the absorption frequency compared with nonlabeled copolymers (See figs. S25 and S26 for the IR spectra of the latter).

Table 1. Polymerization experiments.*

Entry	Cat.	p [bar]	t [min]	CO in feed [mol %]†	Yield [g]	TOF‡	χ§ [mol %]	I / NA / A¶ [mol %]	M_n# [10^3 g mol^{-1}]	M_w/M_n#	T_m [°C] (%cryst.)**
1	**2**	5	10	0.2	0.80	15.6	0.3 (0.4)	69 / 11 / 20	54	1.7	133 (71)
2	**2**	5	60	0.2	2.07	7.41	0.3 (0.4)	70 / 22 / 8	102	1.5	135 (66)
3	**2**	5	120	0.2	4.26	7.61	0.3 (0.3)	79 / 21 / 0	140	1.6	134 (68)
4	**2**	5	120	0.6	1.98	3.53	0.8 (1.1)	32 / 36 / 32	147	1.8	135 (62)
5	**2**	10	240	0.6	3.31	2.95	0.6 (1.0)	43 / 33 / 24	216	1.8	136 (60)
6	**3**	5	120	0.6	1.13	2.02	1.0 (1.4)	21 / 29 / 49	60	1.7	133 (69)
7	**4**	5	120	0.6	0.26	0.46	3.7 (4.4)	28 / 34 / 38	84	1.7	134 (63)
8	**5**	5	120	0.6	0.44	0.79	1.6 (3.0)	7 / 34 / 59	45	1.6	133 (69)
9	**6**	5	120	0.6	0.82	1.47	1.1 (2.5)	18 / 33 /49	90	1.7	135 (67)

*Reaction conditions: 10 μmol precatalyst loading, 100°C, 100 ml of toluene, 500 rpm pitched blade stirrer. †Carbon monoxide content in the feed gas. ‡TOF given in units of 10^3 mol [C_2H_4] mol^{-1} [Ni] h^{-1}. §Carbon monoxide incorporation determined by IR spectroscopy. In brackets: Carbon monoxide incorporation calculated from ^{13}C NMR spectroscopy by integration of the ^{13}C labeled C=O signals in relation to the overall integral, considering natural abundance of ^{13}C versus ^{12}C. ¶I: isolated carbonyls, NA: nonalternating adjacent carbonyls (See Fig. 2 for specification). A, alternating polyketone segments. Determined by ^{13}C NMR spectroscopy. #Determined by GPC in 1,2-dichlorobenzene at 160°C through universal calibration versus polystyrene standards. **Determined by DSC, second heating cycle.

keto content was determined by quantitative NMR spectroscopy and confirmed by IR spectroscopy (Table 1, Fig. 2B, and figs. S25 and S26). As anticipated, these in-chain keto motifs do not compromise the desirable material properties of polyethylene. Wide angle x-ray scattering (WAXS) diffractograms of the materials are virtually identical to that of HDPE (Fig. 3A), underlining a conserved solid-state structure. Also, melting and crystallization points are virtually unaltered (Fig. 3C). Most notably, the keto-modified polyethylenes possess high molecular weights [up to weight-average molecular weight (M_w) ~400,000 g mol^{-1}, M_n ~220,000 g mol^{-1}]. This result is notable in view of the very low molecular weights observed in ethylene-CO copolymerizations yielding any partially nonalternating motifs to date (*22*), and it is a crucial factor underlying useful mechanical properties. Indeed, tensile tests on injection molded test bars showed a ductile behavior comparable to commercial HDPE (Fig. 3D): modulus of elasticity E = 1062 ± 53 MPa and tensile strength σ_y = 26.7 ± 0.2 MPa, compared with literature values of E = 900 MPa and σ_y = 27 MPa for HDPE, and E = 240 MPa and σ_y = 12 MPa for LDPE (*31*). No evidence for undesirable cross-linking that could hamper melt processing was observed; cross-linking promoted by keto motifs has previously been an issue in the thermoplastic processing of commercial polyketone resins (*32*). Exposure of keto-PE films to simulated sunlight confirmed their photodegradability. After irradiation of films floating on a water bath with a light intensity corresponding to ~5 months of natural sunlight in Southern Europe, the onset of degradation was evidenced by embrittlement and mechanical disintegration of the previously ductile samples, and manifested by an observable weight loss of several weight percent. IR spectroscopy on a ^{13}CO-labeled sample showed that in these short-term degradation experiments the keto groups were only partially consumed; in addition, new keto and ester groups formed, which can enable further chain degradation beyond the observed interval. By contrast, a reference sample of keto-free ethylene homopolymer remained unaltered (see figs. S32 to S34 for detailed data).

At given initial reaction conditions, the composition of the resultant polymer did not vary considerably with increasing reaction time or increasing polymer yield (Table 1, entries 1 to 3 and 5 and table S3, entries 15 and 16). A homogeneous material was consistently formed, opposed to perturbation of microstructure and composition as a result of a conceivable alteration of the relative concentrations of the monomers. Compared with ethylene homopolymerization, the copolymerization yields and average activities were lowered (Table 1, entry 1, versus table S4, entry 2; see table S4 for further homopolymerization data). However, this is not because of a problematically rapid catalyst deactivation. Catalysts were stable for hours in CO copolymerizations (Table 1, entries 1 to 5). Presumably, in addition to CO binding, reversible chelating binding of formed keto groups slows down chain growth to a certain extent as concluded from the observation of overall increased productivities with increasing monomer concentration at a given E/CO ratio (Table 1, entries 1 and 4, versus table S3, entries 17 and 16; see fig. S35 for a scheme of reversible deactivation pathways). A threshold polymerization reaction temperature is required to overcome this retardation. Thus, efficient copolymerization with (**2**) requires ca. 100°C, whereas homopolymerizations can proceed rapidly at lower temperatures (table S4, entries 1, 4 to 8). As anticipated, higher polymerization reaction temperatures (>70 °C) also decrease the CO incorporation and in fact are necessary to achieve nonalternation (Table 1, entries 1 to 5, and table S3 entry 17, versus table S3, entry 18). This observation underlines the crucial role of the temperature stability of catalysts (**2**) to (**6**) in enabling the nonalternating copolymerizations of keto-modified polyethylenes. Beyond this, the underlying phosphinophenolate motif also appears particularly well suited to nonalternating copolymerization, with catalyst (**2**) standing out in detail in terms of its combination of productivity, microstructure and produced polymer molecular weight. By comparison, catalyst (**1**)—even at an elevated polymerization temperature—formed only alternating polyketones (table S3, entry 12). This privileged nature of the phosphinophenolate motif may be related to the strong metal-phosphine bond, which can counteract irreversible deactivation by ligand displacement and reduce the preference for CO versus ethylene incorporation by enhancing the electron density at the metal.

The high molecular weights of the keto-modified polyethylenes suggest that the presence and incorporation of CO does not promote any problematic chain transfer reactions. Indeed, an analysis of the polymers' endgroups shows only those types of endgroups also found in homopolyethylenes, namely olefinic endgroups originating from ß-hydride transfer of the Ni-polymeryl species (fig. S16).

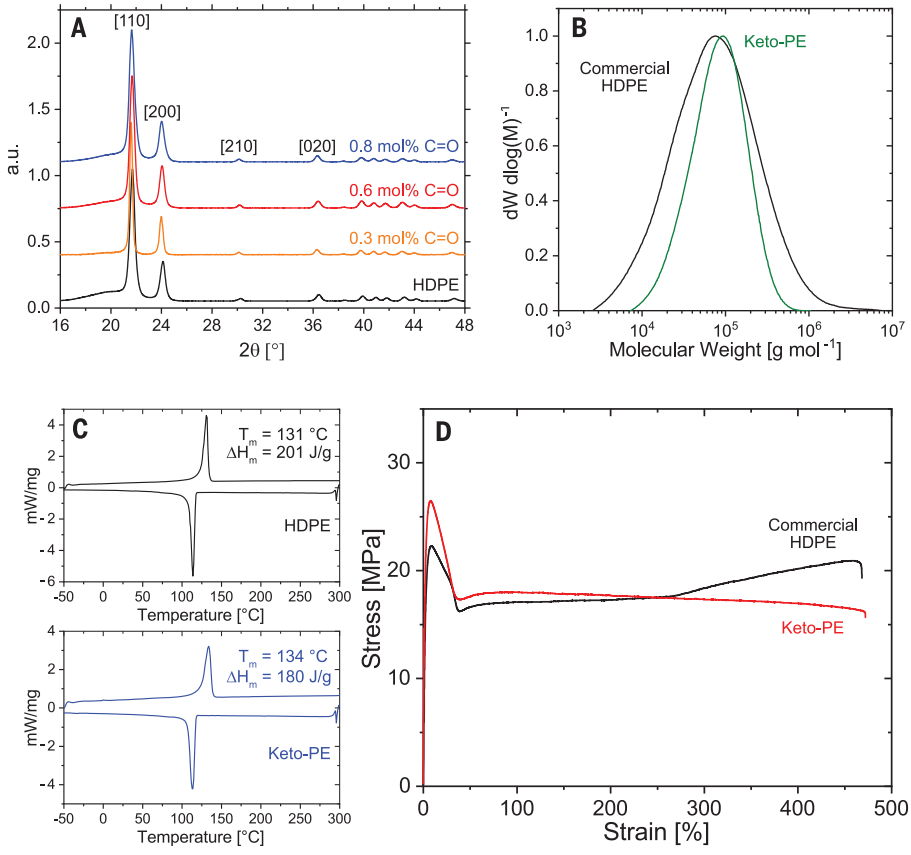


Fig. 3. Materials characterization of copolymers. (**A**) WAXS diffractograms of copolymers with different keto content and commercial HDPE. Reflections correspond to the orthorhombic unit cell of polyethylene (see supplementary materials for further details). (**B**) GPC trace of copolymer (1.0 mol% keto repeat units), and commercial HDPE for comparison. (**C**) DSC trace of copolymer (0.8 mol% keto repeat units), and commercial HDPE for comparison. (**D**) Representative stress-strain curves of injection molded keto-modified polyethylene (0.6 mol% keto repeat units) in comparison to commercial HDPE (see supplementary materials for further data).

This confirms that no undesired additional chain transfer pathways are operative in CO copolymerizations. Rather, the presence of CO appears to hinder chain transfer and branching as concluded from higher molecular weights and the complete absence of branches in copolymers compared with ethylene homopolymers (Table 1, entries 1 to 4, versus table S4, entry 2, fig. S17, and figs. S20 to S22), possibly by blocking of coordination sites for ß-hydride transfer by CO.

Access to in-chain functionalized polyethylene, which possess the desirable material properties of HDPE, offers perspectives for polyolefins that are less persistent in the environment in enabling photolytic and oxidative chain breakdown. The keto contents reported are reasonable in the sense that the lengths of the polyethylene segments between the installed possible break points correspond to alkane lengths that can be mineralized (*33*, *34*). Beyond this rough consideration, the mechanisms and rates of degradation will depend strongly on specific environmental conditions, and in this context an understanding of desirable profiles of the long-term behavior of nonpersistent materials remains to be established (*35*).

REFERENCES AND NOTES

1. M. Stürzel, S. Mihan, R. Mülhaupt, *Chem. Rev.* **116**, 1398 (2016).
2. G. R. Strobl, *The Physics of Polymers: Concepts for Understanding their Structures and Behavior*, (Springer, ed. 3, 2007).
3. T. O. Morgen, M. Baur, I. Göttker-Schnetmann, S. Mecking, *Nat. Commun.* **11**, 3693 (2020).
4. G. H. Hartley, J. E. Guillet, *Macromolecules* **1**, 165 (1968).
5. A. L. Andrady, J. E. Pegram, Y. Song, *J. Environ. Polym. Degrad.* **1**, 117 (1993).
6. P. Ortmann, F. P. Wimmer, S. Mecking, *ACS Macro Lett.* **4**, 704 (2015).
7. A. Bunescu, S. Lee, Q. Li, J. F. Hartwig, *ACS Cent. Sci.* **3**, 895 (2017).
8. L. Chen *et al.*, *Chem* **7**, 137 (2021).
9. F. C. Rix, M. Brookhart, P. S. White, *J. Am. Chem. Soc.* **118**, 4746 (1996).
10. E. Drent, P. H. M. Budzelaar, *Chem. Rev.* **96**, 663 (1996).
11. A. Piovano, J. Zarupski, E. Groppo, *J. Phys. Chem. Lett.* **11**, 5632 (2020).
12. W. Bertleff, M. Roeper, X. Sava, in *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley, 2012), pp. 73–97.
13. Press release by Hyosung of May 5, 2016: Hyosung, introduces the polyketone brand 'POKETONE™' to the global market (retrieved June 8, 2021 from www.poketone.com).
14. Z. Chen, M. Brookhart, *Acc. Chem. Res.* **51**, 1831 (2018).
15. C. Tan, C. Chen, *Angew. Chem. Int. Ed.* **58**, 7192 (2019).
16. A. Keyes *et al.*, *Angew. Chem. Int. Ed.* **58**, 12370 (2019).
17. A. Nakamura, S. Ito, K. Nozaki, *Chem. Rev.* **109**, 5215 (2009).
18. E. Drent, R. van Dijk, R. van Ginkel, B. van Oort, R. I. Pugh, *Chem. Commun.* **9**, 964 (2002).
19. C. Chen *et al.*, *Organometallics* **30**, 5248 (2011).
20. L. Bettucci *et al.*, *Dalton Trans.* **2007**, 5590 (2007).
21. D. K. Newsham, S. Borkar, A. Sen, D. M. Conner, B. Goodall, *Organometallics* **26**, 3636 (2007).
22. S. S. Soomro, D. Cozzula, W. Leitner, H. Vogt, T. E. Müller, *Polym. Chem.* **5**, 3831 (2014).
23. S. Mecking, M. Schnitte, *Acc. Chem. Res.* **53**, 2738 (2020).
24. W. Keim, *Angew. Chem. Int. Ed.* **52**, 12492 (2013).
25. K. A. Ostoja Starzewski, J. Witte, *Angew. Chem. Int. Ed.* **26**, 63 (1987).
26. U. Klabunde, S. D. Ittel, *J. Mol. Catal.* **41**, 123 (1987).
27. W. Keim, F. H. Kowaldt, R. Goddard, C. Krueger, *Angew. Chem. Int. Ed.* **17**, 466 (1978).
28. B. S. Xin *et al.*, *J. Am. Chem. Soc.* **139**, 3611 (2017).
29. Y. Zhang, H. Mu, L. Pan, X. Wang, Y. Li, *ACS Catal.* **8**, 5963 (2018).
30. C. Stafford, B. A. Arndtsen, *Inorg. Chim. Acta* **369**, 231 (2011).
31. D. Jeremic, in *Ullmann's Encyclopedia of Industrial Chemistry*; Wiley, 2014), pp. 1–42.
32. C. E. Ash, D. G. Waters, A. A. Smaardijk, *Soc. Plast. Eng.* **53**, 2319 (1995).
33. J. R. Haines, M. Alexander, *J. Appl. Microbiol.* **28**, 1084 (1974).
34. F. Kawai *et al.*, *Polym. Degrad. Stabil.* **86**, 105 (2004).
35. M. T. Zumstein, R. Narayan, H. E. Kohler, K. McNeill, M. Sander, *Environ. Sci. Technol.* **53**, 9967 (2019).

ACKNOWLEDGMENTS

We thank I. Göttker-Schnetmann for support with NMR data interpretation, N. Mast, M. Eck and M. Häußler for support with polymer synthesis and processing, L. Bolk for DSC and GPC measurements and R. Kirsten for technical support. We further thank D. Dietrich, C. Hauck, and A. Marx and their coworkers for enabling safe laboratory operations by providing rapid COVID testing. **Funding:** The ERC (Advanced Grant DEEPCAT, 832480) is gratefully acknowledged. F.L. thanks the Alexander von Humboldt-Foundation for a postdoctoral research stipend. **Author contributions:** S.M. conceived the concept. M.B., F.L., T.M., L.O., and S.M. jointly devised the experimental programm. F.L., M.B., and T.M. synthesized and characterized catalyst precursors. M.B. and T.M. performed polymerization experiments with phosphinophenolato catalysts. L.O. performed experiments with salicylaldiminato catalysts. M.B. and T.M. performed microstructure analysis and experiments on materials properties. All authors jointly wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. Correspondence and requests for materials should be addressed to S.M.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi8183
Materials and Methods
Figs. S1 to S35
Tables S1 to S5
References (*36*–*47*)

1 April 2021; accepted 5 August 2021
10.1126/science.abi8183

SUPERCONDUCTIVITY

Direct evidence for Cooper pairing without a spectral gap in a disordered superconductor above T_c

Koen M. Bastiaans[1], Damianos Chatzopoulos[1], Jian-Feng Ge[1], Doohee Cho[2], Willem O. Tromp[1], Jan M. van Ruitenbeek[1], Mark H. Fischer[3], Pieter J. de Visser[4], David J. Thoen[5,6], Eduard F. C. Driessen[7], Teunis M. Klapwijk[5,8], Milan P. Allan[1]*

The idea that preformed Cooper pairs could exist in a superconductor at temperatures higher than its zero-resistance critical temperature (T_c) has been explored for unconventional, interfacial, and disordered superconductors, but direct experimental evidence is lacking. We used scanning tunneling noise spectroscopy to show that preformed Cooper pairs exist up to temperatures much higher than T_c in the disordered superconductor titanium nitride by observing an enhancement in the shot noise that is equivalent to a change of the effective charge from one to two electron charges. We further show that the spectroscopic gap fills up rather than closes with increasing temperature. Our results demonstrate the existence of a state above T_c that, much like an ordinary metal, has no (pseudo)gap but carries charge through paired electrons.

The zero-resistance state of superconductivity emerges when electrons form Cooper pairs, which condense into a superfluid with long-range phase coherence. For conventional, elemental superconductors, pairing and condensation take place concurrently when cooling below the critical temperature (T_c). By contrast, disordered superconductors exhibit unusual normal state properties above T_c (*1*–*9*), which, analogous to high-temperature (*10*–*14*), interface (*15*–*17*), and heavy fermion superconductors (*18*, *19*), were thought to be a consequence of preformed or fluctuating Cooper pairs that could exist above T_c. However, there is no direct experimental evidence of this.

There have been several theoretical descriptions for electron pairs without superconductivity, i.e., without phase coherence (Fig. 1A). The best-known models postulate a phase fluctuation–driven breakdown of coherence at T_c (*1*, *20*). Although strictly speaking only valid for neutral superfluids, such a breakdown has a particularly intuitive picture in the Berezinskii-Kosterlitz-Thouless theory: Fluctuating vortex-antivortex pairs that exist below T_c unbind at T_c, leading to the change from superconducting to resistive state as the temperature is increased (*21*). Phase fluctuations are also at the core of models involving the Bose-Einstein condensation (BEC) to Bardeen-Cooper-Schrieffer (BCS) crossover (*22*), which were realized in cold atom ensembles. An alternative scenario for the breakdown of superconductivity is a decrease of the order-parameter amplitude, which is caused by enhanced Coulomb repulsion in disordered systems (*23*). Phase-fluctuation models invoke paired electrons above T_c, whereas pairing amplitude models do not.

The challenge to experimentally distinguishing phase-incoherently paired electrons from single electrons is twofold: (i) many properties of paired but uncondensed electrons are the same as for single electrons, including the charge transported per electron, and (ii) spectroscopic signatures of paired electrons are often similar to single-electron phenomena such as charge density waves (*10*). Nevertheless, there have been many interesting observations connected to pairing above T_c reported in different families of unusual superconductors. First, kinks in resistivity versus temperature curves and deviations from assumed normal state resistances have been connected to pairing fluctuations in disordered and cuprate high-temperature superconductors (*3*, *5*, *11*). Second, the Nernst effect has shown the existence of a pseudogap onset temperature above T_c in cuprate, heavy fermion, and disordered superconductors, which is compatible with short-lived Cooper pair fluctuations (*12*, *13*, *19*). Third, in underdoped cuprates, enhanced noise signatures have been observed in planar junctions and interpreted as pairing of electrons, both in the superconducting state and above T_c (*14*). Enhanced noise has even been observed at bias voltages greater than the superconducting gap. Fourth, several spectroscopic techniques show (partially filled) gaps in the spectral weight at the Fermi level, frequently called pseudogaps, that persist above T_c, in disordered, cuprate, interfacial, and heavy fermion superconductors (*1*, *7*, *10*, *16*, *18*). These observations, which differ in many respects from expectations for the conventional metallic state in ordinary metals, have been interpreted as being due to a finite population of paired electrons.

Our study aimed to determine the nature of the charge carriers in this unconventional normal state in disordered superconductors by focusing on the effective charge of the carriers in tunneling experiments as measured by noise spectroscopy. Shot noise spectroscopy in mesoscopic systems has proven to be a powerful tool to determine the effective charge, e.g., in superconductors or in fractional quantum Hall experiments (*24*). In general, tunneling between two leads biased with voltage V is a Poissonian process. The current noise $S = \langle (I - \langle I \rangle)^2 \rangle$ associated with the granularity of charge is proportional to the effective charge q^* of the carrier and the current I, i.e., $S = 2q^*|I|$. This relation allows the extraction of the effective charge of the carriers, which in metal-insulator-superconductor interfaces (NIS) is equal to one electron charge ($1e$) at biases above the superconducting gap (Fig. 1B) but equal to $2e$ within the gap. The latter is a result of Andreev reflections from paired electrons, which double the effective charge transported (*25*), as illustrated in Fig. 1C. The signature for paired electrons is thus simple and unambiguous: In a tunneling experiment from a normal metal to a system with bound pairs, the normalized noise should change from $S/2I = 1e$ to $S/2I = 2e$ when the bias is reduced to below the gap energy. Experiments involving conventional superconductors have confirmed the doubling, and even further multiplication, of shot noise as a tell-tale signature of paired electrons (*25*–*28*); however, ensuring a clean vacuum barrier has proven to be challenging.

We chose to use the disordered superconductor titanium nitride (TiN) (*29*) for this study because it exhibits robust signatures of unusual physics above T_c without any competing orders such as charge density waves (*10*). Similar to other disordered superconductors, TiN can be tuned toward a superconductor-metal or superconductor-insulator transition (*1*) and exhibits electronic granularity that might be emergent (*30*) or caused by small superconducting islands coupled to each other. Our TiN films developed a zero-resistance state, i.e., became superconducting below 2.95 K as determined by transport measurements, and exhibited a mean free path of 0.57 nm, with a coherence length of ~10 nm, placing them well within the so-called "dirty limit" of superconductivity. We used 45-nm-thick films fabricated by plasma-enhanced atomic layer deposition (ALD) (*29*) and, as control samples, 60-nm-thick films made by sputter deposition on silicon substrates (the data from

[1]Leiden Institute of Physics, Leiden University, 2333 CA Leiden, Netherlands. [2]Department of Physics, Yonsei University, Seoul 03722, Republic of Korea. [3]Department of Physics, University of Zurich, 8057 Zurich, Switzerland. [4]SRON Netherlands Institute for Space Research, 2333 CA Leiden Netherlands. [5]Kavli Institute of Nanoscience, Delft University of Technology, 2628 CJ Delft, Netherlands. [6]Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, Netherlands. [7]Institut de Radioastronomie Millimétrique (IRAM), Grenoble, 38400 Saint-Martin-d'Hères, France. [8]Institute of Topological Materials, Julius-Maximilian-Universität Würzburg, 97070 Würzburg, Germany.
*Corresponding author. Email: allan@physics.leidenuniv.nl

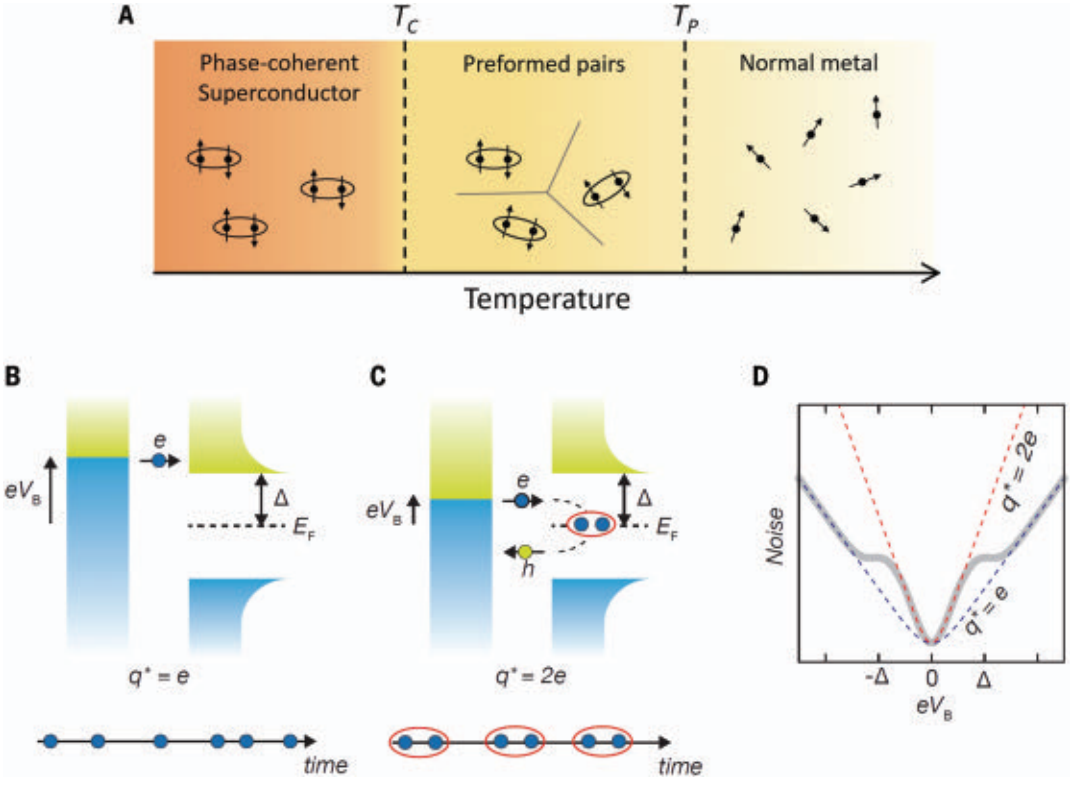


Fig. 1. Noise spectroscopy as a direct probe to detect paired electrons. (**A**) Illustration of the different electronic states. At high temperature, a conventional metal state consists of single electrons. Below T_c, these electrons couple to form a phase-coherent state of Cooper pairs. Between these two regimes, an additional state of non-phase-coherent, preformed Cooper pairs is conjectured to exist. (**B**) "Normal" NIS transport of single electron charge. The characteristic density of states of the superconducting sample is shown, with filled and empty states denoted by blue and yellow, respectively, separated by a pair-breaking gap Δ. (**C**) Andreev reflection process in a BCS superconductor. An electron transfers a Cooper pair into the superconductor by reflecting a hole in the opposite direction, effectively transferring 2*e* charge. (**D**) Illustration of noise as a function of bias voltage for $q^* = 1e$ and $q^* = 2e$ transport. For an NIS junction, the expected noise is indicated by the gray curve.

the sputtered sample were qualitatively similar; figs. S4, C and D, S6, and S7). The samples were inserted into an ultrahigh vacuum chamber and then into a cryogenic scanning tunneling microscope (STM). The noise measurements were done with a custom-built, cryogenic MHz amplifier consisting of a superconducting tank circuit and custom-built high electron mobility transistors, as described elsewhere (*31*). From the noise spectroscopic measurements, we measured the current fluctuations around a center frequency of 3 MHz to avoid mechanical resonances and unwanted 1/*f* noise and ensured that the vacuum tunneling barrier was clean by repeatedly measuring topographies with the STM (*32*) (fig. S1). Spectroscopic imaging STM at 2.3 K revealed a partially filled gap of Δ = 1.78 meV (with spatial variations of ~36% (fig. S4).

Local tunneling noise spectroscopy is the key technique used in this study. We performed our experiments at fixed junction resistance (R_J), so the noise is expected to be proportional to the bias voltage, $S(q^*,V) = 2q^*|I| = 2q^*|V/R_J|$. At finite temperature (T) and low junction transmission, the formula is modified to $S(q^*,V) = 2q^*(V/R_J)\coth(q^*V/2k_BT)$, where k_B is the Boltzmann constant (*25*). We extracted the effective charge q^* by numerically solving this formula for the observed shot noise at each bias (*33*). As expected, the effective charge at a bias higher than Δ was equal to one electron charge, $q^* = 1e$, as shown in Fig. 2, A and B. However, a clear change in the effective charge from $q^* = 1e$ to $q^* = 2e$ is visible in the data for voltages below gap energy. This is unambiguous evidence that the electrons in TiN films were paired below an energy of ~Δ (indicated in Fig. 2 by blue shading). The reason that the noise did not rise immediately at Δ but at energies just below Δ is thermal broadening (*33*) The shape and values of our noise spectra enabled us to directly deduce pairing as the source of the noise, as opposed to fluctuating orders that might be present in the sample.

The noise enhancement to 2*e* persisted when warming the sample to temperatures above the zero-resistance T_c. Figure 3 shows noise spectra acquired at different temperatures ranging from 2.3 to 7.2 K, which correspond to 0.78 and $2.43T_c$, respectively. Up to more than twice T_c, the noise spectra still show enhanced noise corresponding to 2*e*; only at $T = 2.43T_c$ does the noise decrease below 2*e*. Given that T_c is far below the temperature at which the noise is enhanced, there is another transition temperature, which we denote here by T_p, associated with pairing. In a fluctuation picture, this would be the temperature at which the gap opens. Figure 4 summarizes the temperature evolution of the noise.

We can compare how T_p scales with the unusual transport properties that have been analyzed for a wide range of disordered superconductors. Figure 4C shows a resistance versus temperature curve showing the superconducting transition at $T_c = 2.95$ K. At ~11 K, the resistance curve shows a so-called *N*-shaped curvature (fig. S3), as is typical for disordered superconductors not too close to the superconductor-insulator transition (*3*) and for cuprate high-temperature superconductors (*34*). This local maximum at 11 K is a signature that has been interpreted as the onset of superconducting fluctuations (*3*, *35*). For the ALD sample in Fig. 4C, this is also roughly the temperature at which the gap is expected to close if the ratio $2\Delta(0)/k_BT_p$ is given by the BCS value of 3.52; for the sputtered sample, we obtained a lower mean-field prediction for T_p (fig. S7).

Our most unexpected experimental observation was the pairing above T_c even in the absence of a spectroscopic (pseudo)gap. As shown in Fig. 4, A and B, the gap in the differential tunneling conductance of TiN filled up at T_c when increasing the temperature instead of closing, i.e., $\Delta(T)$ was constant while the spectral weight inside the gap was filling up. To ensure that the 2*e* noise was not incidentally measured in coherent superconducting "puddles," we measured these

spectra at the same locations where we performed the noise measurements (figs. S1 and S4). Although the measured temperature evolution $\Delta(T)$ resembles findings in other disordered superconductors, the gap filled faster in our samples upon increasing temperature such that the gap was fully filled at T_c. Partial gap filling has been observed with various probes in other disordered and unconventional superconductors (*7*, *36*–*38*). The gap filling can be calculated within models involving strong fluctuations of the order parameter (*39*) or large level spacing in grains (*40*); alternatively, one can postulate a large fraction of unpaired electrons, or electrons with very small superconducting gaps, to exist in parallel to the superfluid (*36*, *41*). Our data show that the current noise continued to correspond to ~2*e* at elevated temperatures despite the filling of the gap. The state above T_c thus behaves like an ordinary metal from a spectroscopy point of view but with tunneling current fluctuations that indicate pairing. This is only visible in shot noise experiments. Therefore, a putative coexistence of paired and unpaired electrons, as predicted to exist by theories of short-lived Cooper pairs (*1*, *12*, *13*), is not present in TiN.

We note that the combined observation of a filled gap and 2*e* noise cannot be described in the same way as the well-known case of subgap current in break junctions of elemental superconductors (*25*). In break junctions with low transparencies, the conductance inside the gap is much smaller than outside the gap because Andreev reflections happen with a probability proportional to the square of the transparency of the junction (t^2), whereas single-electron tunneling outside the gap occurs with probability t. By contrast, in TiN, we measured a similar conductance inside

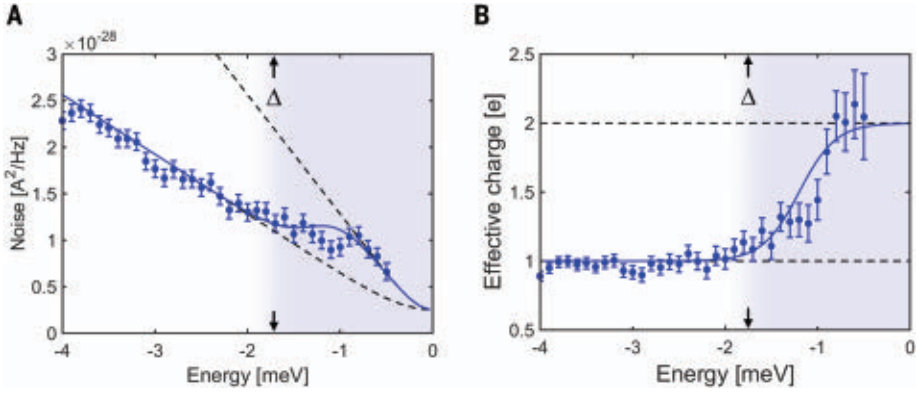


Fig. 2. Evidence for pairing in TiN from scanning noise spectroscopy. (**A**) Noise in the tunneling junction (R_J = 5 MΩ) between the STM tip and TiN sample at 2.3 K, with the thermal amplifier noise subtracted, as function of the bias voltage. (**B**) Spectroscopy of the effective charge $q^*(V)$ at 2.3 K. In both panels, blue points represent experimental data and dashed lines indicate the expected noise for q^* = 1*e* and q^* = 2*e*. Blue shading indicates the spectral gap observed in the differential conductance (Fig. 4A). The solid blue curves indicate a step from 1*e* to 2*e*, broadened by the respective thermal resolution (fig. S8).

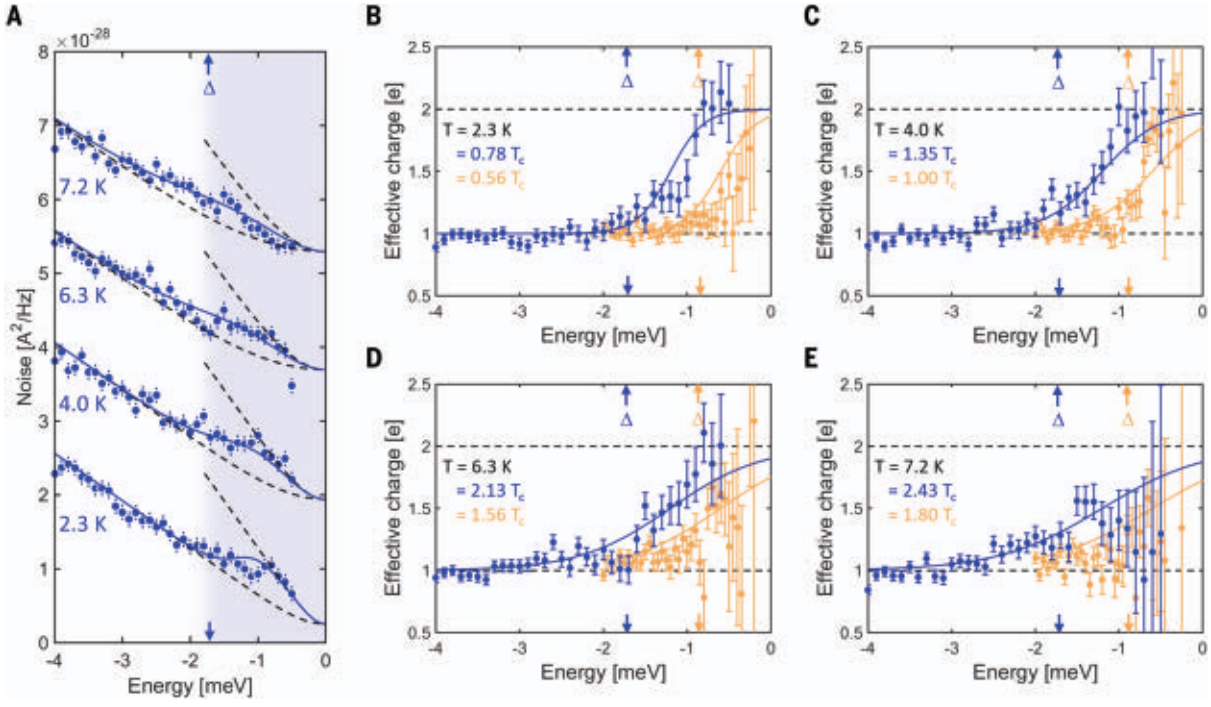


Fig. 3. Enhanced noise above T_c. (**A**) Noise spectroscopy on a TiN sample for different temperatures from 2.3 K = 0.78 T_c to 7.2 K = 2.43 T_c. Blue points indicate the measured excess noise in the junction (R_J = 5 MΩ) as function of bias voltage. The different temperature curves are offset for clarity. Dashed lines indicate the expected noise for q^* = 1*e* and q^* = 2*e*. Solid lines indicate the expected noise when a change from 1*e* to 2*e* occurs at Δ, including the thermal resolution at the given temperature. Blue shading highlights the spectral gap in the differential conductance. (**B** to **E**) Effective charge $q^*(V)$ for the four different temperatures. Data for the ALD (sputtered) sample are shown in blue (orange). The solid blue and orange curves indicate a step from 1*e* to 2*e* at Δ, broadened by the respective thermal resolution (fig. S8).

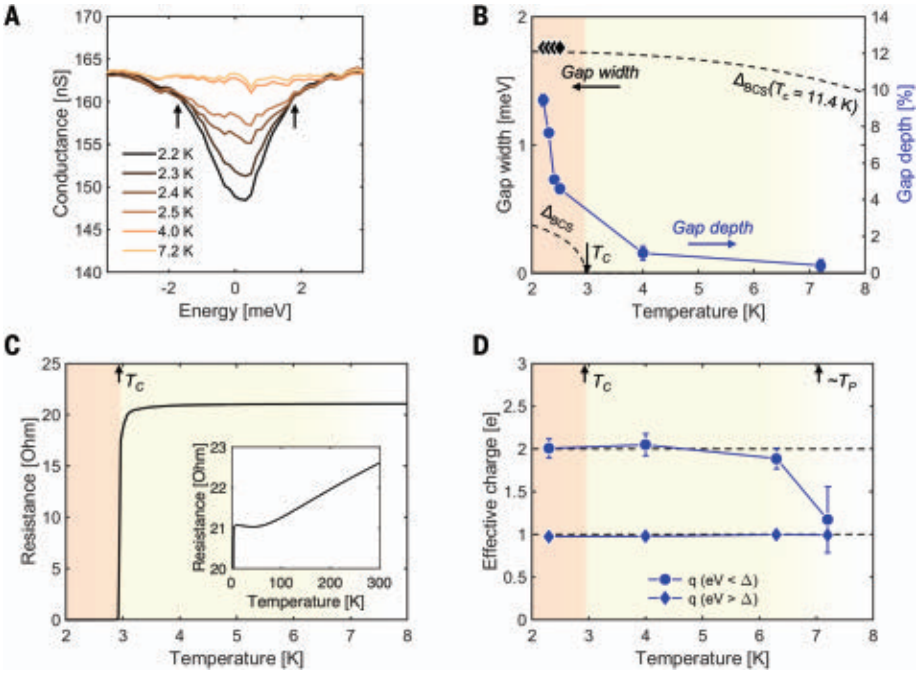


Fig. 4. Evidence for a preformed-pair phase above T_c. Shown are the data for the ALD TiN sample. Data for the sputtered samples can be found in fig. S7. (**A**) Temperature dependence of the spectral density gap measured by the differential tunneling conductance between 2.2 K (0.74 T_c) and 7.2 K (2.43 T_c). Black arrows indicate the gap width at 2.2 K determined by finding the minimum of the second derivative. Setup conditions: V_{bias} = 5 mV, I_{set} = 1 nA. (**B**) Gap width (black diamonds) as a function of temperature for the curves in (A). The dashed curves indicate the mean-field predictions given the resistive T_c, $\Delta(T_c = 2.96\ K)$, and the T_c from fig. S3, $\Delta(T_c = 11.4\ K)$, from BCS theory. The depth of the gap at zero bias (blue dots) for the curves in (A) is shown in percentages with respect to the conductance at energies outside the gap. Conventional Andreev processes or thermal broadening cannot account for the filling observed here (*33*). (**C**) Resistance versus temperature curve of the ALD TiN sample. The orange-shaded region indicates the phase-coherent superconducting phase below the transition temperature. Inset shows the resistance-temperature relation up to 300 K. (**D**) Effective charge outside (diamonds) and inside (circles) the spectral gap as a function of temperature. The region consisting of preformed pairs includes temperatures at which the gap is fully filled and is indicated by yellow shading.

and outside that gap that was independent of the junction resistance, despite the fact that our transparency was $\sim 2.6 \times 10^{-3} \ll 1$. The probability for charge transfer in the range of $2e$ noise is therefore still linear in transparency. Such a situation could, in principle, arise when the bunching of the probability for subsequent electron transfers is modulated because of Andreev processes within the sample or the diffusive character of the charge in disordered metals (*25*). More likely, a theory involving the spatial heterogeneity and correlations that are typical for disordered superconductors is needed to understand this peculiar state.

In summary, we have used local noise spectroscopy as an unambiguous probe of pairing in a disordered superconductor. We have shown that (i) pairing dominates up to a temperature scale T_p much larger than T_c, (ii) the energy of pairing is related to the gap energy, and (iii) even though the spectral gap is partly or fully filled, almost all observed electrons are paired, differentiating between proposals for pairing above T_c. Therefore, we have observed a state that exhibits $2e$ noise despite having the characteristics of an ordinary metal in differential conductance, without a spectroscopic (pseudo)gap. Further, our results contradict theories of the breakdown of superconductivity that involve a large fraction of unpaired electrons.

REFERENCES AND NOTES

1. B. Sacépé, M. Feigel'man, T. M. Klapwijk, *Nat. Phys.* **16**, 734–746 (2020).
2. N. Hadacek, M. Sanquer, J. C. Villégier, *Phys. Rev. B* **69**, 024505 (2004).
3. S. V. Postolova, A. Y. Mironov, M. R. Baklanov, V. M. Vinokur, T. I. Baturina, *Sci. Rep.* **7**, 1718 (2017).
4. C. Carbillet *et al.*, *Phys. Rev. B* **93**, 144509 (2016).
5. T. I. Baturina *et al.*, *JETP Lett.* **79**, 337–341 (2004).
6. D. B. Haviland, Y. Liu, A. M. Goldman, *Phys. Rev. Lett.* **62**, 2180–2183 (1989).
7. B. Sacépé *et al.*, *Nat. Commun.* **1**, 140 (2010).
8. T. Dubouchet *et al.*, *Nat. Phys.* **15**, 233–236 (2019).
9. P. Brighi, M. Grilli, B. Leridon, S. Caprara, *Phys. Rev. B* **100**, 174518 (2019).
10. B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, J. Zaanen, *Nature* **518**, 179–186 (2015).
11. P. M. C. Rourke *et al.*, *Nat. Phys.* **7**, 455–458 (2011).
12. K. Behnia, H. Aubin, *Rep. Prog. Phys.* **79**, 046502 (2016).
13. O. Cyr-Choinière *et al.*, *Phys. Rev. B* **97**, 064502 (2018).
14. P. Zhou *et al.*, *Nature* **572**, 493–496 (2019).
15. G. Cheng *et al.*, *Nature* **521**, 196–199 (2015).
16. C. Richter *et al.*, *Nature* **502**, 528–531 (2013).
17. I. Božović, J. Levy, *Nat. Phys.* **16**, 712–717 (2020).
18. K. Shrestha *et al.*, *Phys. Rev. B* **103**, 224515 (2021).
19. T. Yamashita *et al.*, *Nat. Phys.* **11**, 17–20 (2015).
20. M. P. A. Fisher, *Phys. Rev. Lett.* **65**, 923–926 (1990).
21. J. M. Kosterlitz, *Rep. Prog. Phys.* **79**, 026001 (2016).
22. Y. L. Loh, M. Randeria, N. Trivedi, C. Chang, R. Scalettar, *Phys. Rev. X* **6**, 021029 (2016).
23. A. M. Finkel'stein, *Phys. B* **197**, 636–648 (1994).
24. R. de-Picciotto *et al.*, *Nature* **389**, 162–164 (1997).
25. Ya. M. Blanter, M. Büttiker, *Phys. Rep.* **336**, 1–166 (2000).
26. X. Jehl, M. Sanquer, R. Calemczuk, D. Mailly, *Nature* **405**, 50–53 (2000).
27. P. Dieleman, H. G. Bukkems, T. M. Klapwijk, M. Schicke, K. H. Gundlach, *Phys. Rev. Lett.* **79**, 3486–3489 (1997).
28. K. M. Bastiaans *et al.*, *Phys. Rev. B* **100**, 104506 (2019).
29. E. F. C. Driessen, P. C. J. J. Coumou, R. R. Tromp, P. J. de Visser, T. M. Klapwijk, *Phys. Rev. Lett.* **109**, 107003 (2012).
30. K. Bouadim, Y. L. Loh, M. Randeria, N. Trivedi, *Nat. Phys.* **7**, 884–889 (2011).
31. K. M. Bastiaans *et al.*, *Rev. Sci. Instrum.* **89**, 093709 (2018).
32. W. T. Liao *et al.*, *Phys. Rev. B* **100**, 214505 (2019).
33. Materials and methods are available as supplementary materials.
34. R. Daou *et al.*, *Nat. Phys.* **5**, 31–34 (2009).
35. T. I. Baturina *et al.*, *EPL* **97**, 17012 (2012).
36. M. A. Tanatar *et al.*, *Phys. Rev. Lett.* **95**, 067002 (2005).
37. T. J. Reber *et al.*, *Phys. Rev. B* **87**, 060506 (2013).
38. M. C. Boyer *et al.*, *Nat. Phys.* **3**, 802–806 (2007).
39. A. V. Chubukov, A. Abanov, I. Esterlis, S. A. Kivelson, *Ann. Phys.* **417**, 168190 (2020).
40. J. von Delft, *Ann. Phys.* **10**, 219–276 (2001).
41. J. S. Van Dyke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11663–11667 (2014).
42. Data for: K. M. Bastiaans, D. Chatzopoulos, J.-F. Ge, D. Cho, W. O. Tromp, J. M. van Ruitenbeek, M. H. Fischer, P. J. de Visser, D. J. Thoen, E. F. C. Driessen, T. M. Klapwijk, M. P. Allan, Direct evidence for Cooper pairing without a spectral gap in a disordered superconductor above Tc, Zenodo (2021); https://doi.org/10.5281/zenodo.5155148.

ACKNOWLEDGMENTS

We thank C. W. J. Beenakker, T. Benschop, G. Blatter, J. C. Davis, R. Fermin, J. Jesudasan, T. Mechielsen, K. van Oosten, P. Raychaudhuri, B. Sacépé, D. Scholma, N. Trivedi, G. Verdoes, and J. Zaanen for valuable discussions. **Funding:** This work was supported by the European Research Council (ERC StG SpinMelt) and by the Netherlands Organization for Scientific Research (NWO/OCW) as part of the Frontiers of Nanoscience program, as well as through a Vidi grant (680-47-536). D.Cho. was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (grants 2020R1C1C1007895 and 2017R1A5A1014862) and the Yonsei University Research Fund (grant 2019-22-0209). P.J.dV. was supported by the Netherlands Organization for Scientific Research NWO (Veni grant 639.041.750); **Author contributions:** K.M.B, D.Cha., J.-F.G., D.Cho, and W.O.T. performed the experiments and analyzed the data. P.J.dV., D.J.T., E.F.C.D., and T.M.K. fabricated and characterized the samples. All authors contributed to the interpretation of the data and writing of the manuscript. M.P.A. supervised the project. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data presented in this paper are publicly available through Zenodo (*42*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abe3987
Materials and Methods
Figs. S1 to S8
References (*43*–*47*)

19 August 2020; accepted 14 September 2021
10.1126/science.abe3987

PLASMONICS

Electrically switchable metallic polymer nanoantennas

Julian Karst[1], Moritz Floess[1], Monika Ubl[1], Carsten Dingler[2], Claudia Malacrida[2], Tobias Steinle[1], Sabine Ludwigs[2], Mario Hentschel[1], Harald Giessen[1]*

Electrical switching of a metal-to-insulator transition would provide a building block for integrated electro-optically active plasmonics. In this work, we realize plasmonic nanoantennas from metallic polymers, which show well-pronounced localized plasmon resonances in their metallic state. As a result of the electrochemically driven optical metal-to-insulator transition of the polymer, the plasmonic resonances can be electrically switched fully off and back on at video-rate frequencies of up to 30 hertz by applying alternating voltages of only ±1 volt. With the use of this concept, we demonstrate electrically switchable beam-steering metasurfaces with a 100% contrast ratio in transmission. Our approach will help to realize ultrahigh efficiency plasmonic-based integrated active optical devices, including high-resolution augmented and virtual reality technologies.

Miniaturization of optical components is key for achieving ultimate spatiotemporal control of light, which will in turn enable researchers to facilitate and improve emerging optical technologies such as augmented and virtual reality, dynamic holography, LIDAR (light detection and ranging), and high-resolution wavefront and polarization shapers. Subwavelength nanoantennas are a prime candidate to achieve this goal (*1*–*5*). Their resonant nature allows highly efficient interaction with light and can locally enhance the manipulation, detection, or modulation of light on ultrasmall, subwavelength scales. The advent of metasurfaces has provided tools with which to realize plasmon-based spatially dependent static light control on nanometer scales (*6*–*8*). These include ultrathin optical elements, such as metalenses (*9*, *10*), made from metals as well as dielectrics for generalized amplitude and phase control (*11*, *12*). Integration into dynamic electro-optical components requires active spatiotemporal light control with nanoantennas via a dynamic change of the optical properties upon an external stimulus (*13*, *14*). One possibility is to operate nanoantennas close to a phase transition, thus resulting in switchable plasmonic resonances and active metasurfaces with high contrast ratios (*15*, *16*). Such direct modifications rely on structural or chemical material transitions stimulated by chemical reactions (*15*), gas exposure (*17*, *18*), or temperature (*19*, *20*). They are usually slow, with incomplete reversibility, and pose major challenges for integration with electro-optical devices. One alternative is an electrically or temperature-driven change of the surrounding medium, tuning the resonance intensity, position, and linewidth, properties which limit the achievable contrast in active metasurface applications (*21*–*29*). Yet, a nanoantenna resonance that can be turned fully ON and OFF electrically at high frequencies has not been achieved to date.

Here we introduce such electrically switchable nanoantennas, made from an optically metallic polymer with an electrochemically driven metal-to-insulator transition in the near–infrared (IR) spectral range caused by a variation of the charge carrier density. In our electrically switchable plasmonic system, the nanoantennas are made from a metallic polymer, which can be electrically switched via the applied voltage (Fig. 1A). When a voltage of +1 V is applied, the polymer is electrochemically doped and oxidized, which results in high carrier density and metallic optical properties. Consequently, the polymer nanoantenna is turned ON and exhibits a strong plasmonic nanoantenna resonance. In contrast, at an applied voltage of −1 V, the carrier density is substantially reduced and the polymer becomes insulating. The nanoantenna is switched OFF,

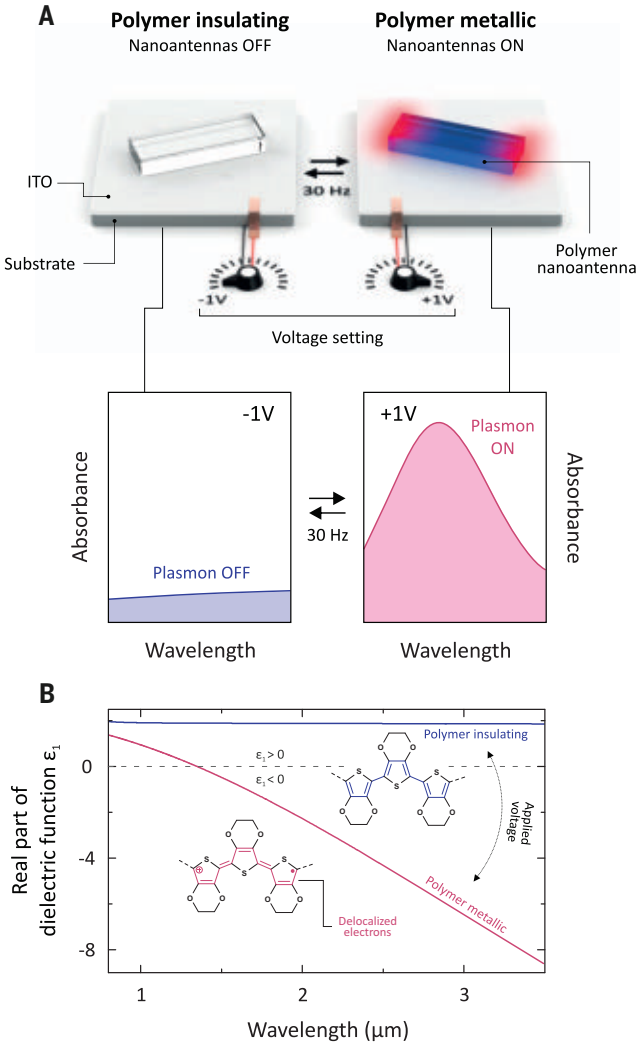


Fig. 1. Concept of electrically switchable nanoantennas made from metallic polymers with a metal-to-insulator transition. (**A**) Electrical switching of an individual metallic polymer nanoantenna. (Right) Antenna ON state at an applied voltage of +1 V. The polymer is in the metallic state and the nanoantenna exhibits a plasmonic resonance. (Left) Antenna OFF state at −1 V. The polymer becomes insulating and the nanoantenna exhibits no plasmonic resonance. Switching between the states reaches video-rate frequencies of 30 Hz. ITO (indium tin oxide) is used for electrical contacting. (**B**) Real part of the dielectric function ε_1 of the polymer in the metallic (red) and insulating (blue) states. A metallic polymer with $\varepsilon_1 < 0$ is obtained in the near-IR and mid-IR for wavelengths $\lambda > 1.3$ μm. By means of the applied voltage, the polymer can be switched into an insulating state with $\varepsilon_1 > 0$ in the entire wavelength range. (Insets) Structural conversion of the metal-to-insulator transition.

[1]4th Physics Institute and Research Center SCoPE, University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany. [2]IPOC-Functional Polymers, Institute of Polymer Chemistry and Center for Integrated Quantum Science and Technology (IQST), University of Stuttgart, Pfaffenwaldring 55, 70569 Stuttgart, Germany.
*Corresponding author. Email: giessen@physik.uni-stuttgart.de

and no plasmonic resonance is observed. Electrical switching between the ON and OFF states occurs very rapidly, permitting video-rate switching frequencies of 30 Hz.

The core of our concept is an electrochemically driven metal-to-insulator transition of the polymer poly(3,4-ethylenedioxythiophene):polystyrene sulfonate (PEDOT:PSS). Figure 1B depicts the real part of the dielectric function ε_1 of PEDOT:PSS in its metallic (red) and insulating (blue) states, indicating this material's excellent electrical and optical properties (see fig. S1 for details). The insets of Fig. 1B illustrate the redox reaction that occurs when PEDOT:PSS changes from its neutral (insulating) state to its oxidized (bi-)polaronic (metallic) state. In detail, PEDOT:PSS is alternately oxidized and reduced, as is common in electrochemistry. Thus, applying voltages of +1 V and −1 V (versus a reference electrode) causes doping and dedoping, respectively, of the polymer and alters the charge carrier density while some hysteresis is present (see materials and methods). Higher voltages (outside of our electrochemical potential window) will cause overdoping and degradation of the material. We find that PEDOT:PSS is switchable between metallic ($\varepsilon_1 < 0$) and insulating ($\varepsilon_1 > 0$) states via the applied voltage for wavelengths >1.3 μm. The crossing $\varepsilon_1 = 0$ at $\lambda_p = 1.3$ μm defines the plasma frequency

$$\omega_p = \sqrt{ne^2/\epsilon_0 m_e}$$

below which the material is optically metallic, allowing for localized plasmonic resonances. Modifying ω_p for metallic polymers is therefore possible through manipulation of the (quasi-)free charge carrier density n (e, electron charge; ϵ_0, permittivity; m_e, electron mass). Thus, ultrahigh doping levels in metallic polymers will allow, in the future, their plasma frequency to be pushed into the visible wavelength range. Commercially available PEDOT:PSS in its pristine spin-coated state exhibits a dielectric function nearly identical to that of the oxidized metallic state. Thus, even without any applied voltage, the plasmonic properties of the metallic polymer are fully accessible. Figure S2 illustrates the fabrication and resonance tunability via geometry.

Electrical switching is carried out in a liquid environment using an electrochemical cell with a three-electrode setup (Fig. 2A; see materials and methods and fig. S5A for further information). We measured the spectral response of an array of polymer nanoantennas with length L = 300 nm for different applied voltages against the reference electrode (Fig. 2B). The pristine spectral response (dry state) and the response for +1 V are shown in gray and red, respectively. The nanoantennas exhibit a plasmonic resonance around ~2.4 μm in the dry state, as the PEDOT:PSS is almost completely oxidized and thus metallic. Applying +1 V triggers a further oxidation of PEDOT:PSS with a

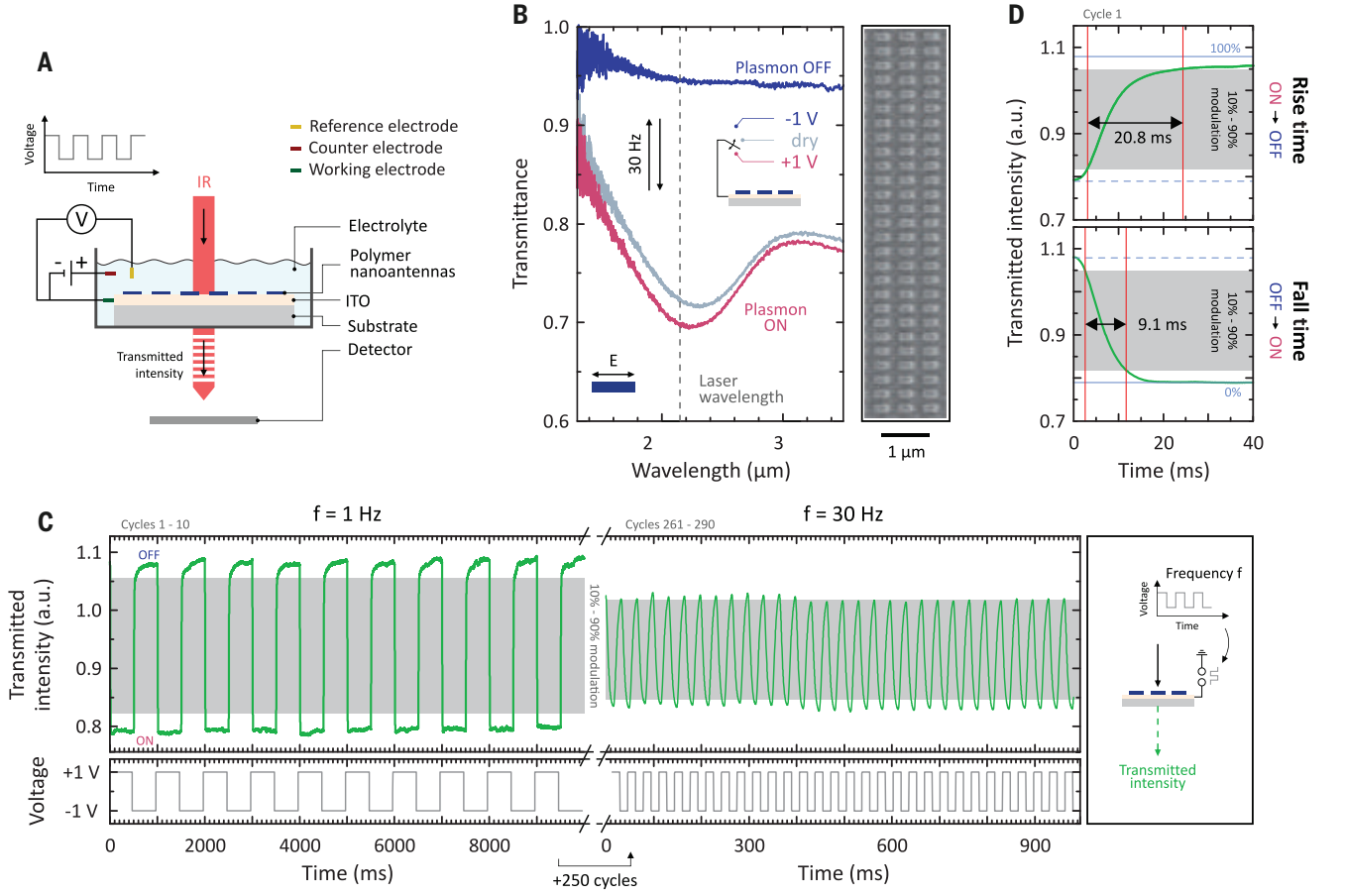


Fig. 2. Video-rate electrical switching of metallic polymer nanoantennas. (**A**) Schematic of the electrochemical cell (three-electrode setup) and optical measurement setup. (**B**) SEM image and spectral response of metallic polymer nanoantennas (length L = 300 nm, width W = 160 nm, height H = 90 nm, periodicity in x P_x = 500 nm, periodicity in y P_y = 300 nm) for different states. Dry state, pristine (gray): Plasmonic resonance almost completely turned ON. +1 V (red): Plasmonic resonance completely turned ON with highest modulation (polymer fully metallic). −1 V (blue): Plasmonic resonance turned OFF (polymer insulating). Electric field E polarized parallel to the nanoantenna long axis. The dashed line indicates the laser wavelength λ = 2.15 μm for (C) and (D). (**C**) (Top) Transmitted intensity through the polymer nanoantennas cycling between the ON and OFF states. The gray area denotes the 10% to 90% modulation window. (Bottom) Set voltage switching between +1 V and −1 V. (Left) Cycles 1 to 10 with switching frequency f = 1 Hz. (Right) Cycles 261 to 290 with f = 30 Hz. a.u., arbitrary units. (**D**) Detailed analysis of rise time τ_{rise} = 20.8 ms (ON to OFF) and fall time τ_{fall} = 9.1 ms (OFF to ON) for the first switching cycle.

maximum doping level and charge carrier density of the polymer due to the existence of (bi-)polarons (*30*, *31*). The plasmon resonance is fully turned ON with the highest possible modulation (red curve). The peak wavelength shifts slightly blue to ~2.2 μm. Notably, applying a negative voltage of −1 V (blue curve) turns the antenna resonance entirely OFF, with no remaining resonant light interaction. The PEDOT:PSS is completely reduced (i.e., in the neutral state) and hence is dielectric in the studied IR spectral range.

The switching speed of our metallic polymer nanoantennas is investigated in Fig. 2C. The antennas are switched between ON and OFF states at applied voltages of +1 V and −1 V while the transmitted intensity is recorded. The bottom graphs (gray curves) show the voltage applied to the polymer nanoantennas (±1 V), and the corresponding modulated transmitted intensity is depicted in the top graphs (green curves). The left graphs in Fig. 2C illustrate the first 10 switching cycles at a frequency of $f = 1$ Hz, confirming full transitions between the

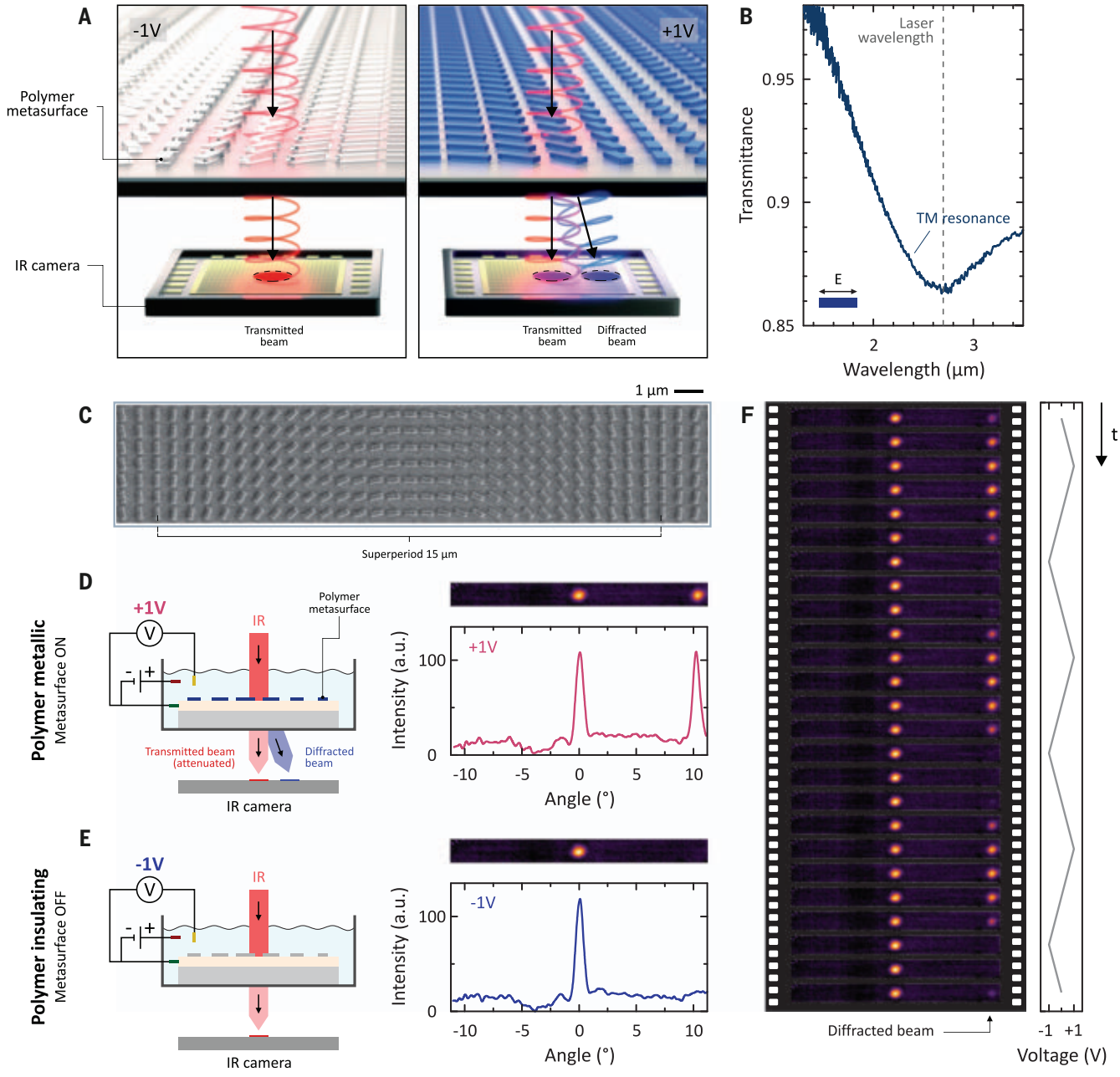


Fig. 3. Metallic polymer plasmonic metasurface for ultrahigh-contrast active beam steering. (**A**) The metallic polymer metasurface is switchable between the OFF state (−1 V, left) and the ON state (+1 V, right), thus allowing beam steering with a 100% contrast ratio in the diffracted beam. (**B** and **C**) Spectrum and SEM image of the pristine metasurface with a superperiod of 15 μm (diffraction angle $\phi = 10.2°$, L = 380 nm, W = 160 nm, H = 90 nm, $P_{x,y}$ = 500 nm). TM (transverse magnetic) plasmonic nanoantenna resonance peaking at 2.65 μm [laser wavelength in (D) to (F)]. (**D** and **E**) (Left) Schematics of the measurement setup and beam-steering concept. The metasurface is turned ON and OFF while the transmitted and diffracted beams are recorded using an IR camera. (Right) Camera images and intensity profiles in the (D) ON and (E) OFF states. The diffracted beam (right peak) at +10.2° vanishes completely for −1 V. (**F**) Selected IR camera images showing three cycles of active beam steering when the applied voltage is cycled in situ between −1 V and +1 V. The intensity of the diffracted beam (right spots) can be continuously varied depending on the applied voltage. t, time.

ON and OFF states without noticeable degradation (gray area: 10% to 90% modulation window; see fig. S5B for further information). A zoomed-in view of the first cycle is displayed in Fig. 2D. We analyze the rise time from ON to OFF (top graph) as well as the fall time from OFF to ON (bottom graph). The rise (or fall) time is defined as the time step in which the intensity rises (or falls) between 10% and 90%, respectively (red in Fig. 2D). We obtain a rise time (τ_{rise}) of 20.8 ms and a fall time (τ_{fall}) of 9.1 ms, and thus a duty cycle time (τ) of 29.9 ms, equivalent to a maximum switching frequency of f = 33 Hz. Thirty cycles of electrical switching of our plasmonic polymer nanoantennas with video-rate frequency of f = 30 Hz are depicted in the right graph of Fig. 2C (gray area: 10% to 90% modulation window, accounting for degradation of nanoantennas after 260 cycles). We find that electrical switching at video-rate frequencies is possible, as the transmitted intensity reaches beyond the 10% to 90% modulation window. The degradation after 290 switching cycles is on the order of only 25%, reducing the overall modulation (comparison of gray marked areas). Possible sources for degradation might be volume expansion of the polymer during switching and irreversible reactions during the electrochemical oxidation and reduction.

Our concept greatly boosts the integrability of plasmonic systems into, e.g., commercial smart and small-scale electro-optical devices, owing to the high switching modulation efficiency with full ON and OFF states, electrical switchability, low required voltages, and switching at video-rate frequencies. Plasmonic metasurfaces are one archetype integration of plasmonics into functional devices. As a proof of concept, we thus demonstrate an electrically switchable metallic polymer metasurface for ultrahigh-contrast active beam steering—that is, a metasurface with the ability to actively control the routing of incident light into a fixed angular range. The basic working principle is illustrated in Fig. 3A: The metasurface is illuminated with a circularly polarized light beam. Depending on the state of the polymer nanoantennas, part of the incident light is diffracted, showing opposite handedness. The key feature of our metallic polymer metasurface is the ability to fully switch it ON and OFF electrically. Consequently, the contrast ratio, defined as the ratio of the diffracted intensities in the metasurface ON and OFF states, reaches 100%.

The TM (transverse magnetic) resonance spectrum of the metallic polymer nanoantennas (the building blocks of the metasurface) peaks at 2.65 μm (Fig. 3B). A scanning electron microscopy (SEM) image of the polymer metasurface is shown in Fig. 3C. The experimental scheme is illustrated on the left in Fig. 3, D and E. The laser (right circularly polarized) is set to an illumination wavelength of 2.65 μm, and the IR camera is used to image the transmitted and diffracted intensities. See fig. S6 for further spectral and imaging information. The measured IR camera images and intensity profiles for an applied voltage of +1 V and −1 V are displayed on the right of Fig. 3, D and E (the primary transmitted beam is attenuated to prevent saturation of the IR camera). A voltage of +1 V turns the metasurface ON, and diffraction by the plasmonic polymer metasurface is observed (Fig. 3D, right). In contrast, the applied voltage of −1 V turns the metasurface completely OFF, and the diffracted beam at +10.2° vanishes completely (Fig. 3E, right). The diffraction efficiency is 1.1%. So far, the spectral contrast of the beam-steering metasurface in its ON state accounts for roughly 86% transmission versus 100% in the OFF state (Fig. 3B). This contrast is currently limiting the diffraction efficiency. Increasing the modulation depth by optimizing material, geometry, and doping levels will enhance the diffraction efficiency.

The metallic polymer nanoantennas possess another intriguing property: Successive electrochemical doping results in intermediate states between the ON and OFF states; thus, the intensity of the diffracted beam can be modified at will (Fig. 3F). We show selected IR camera images for three full switching cycles (voltage cycled between +1 V and −1 V at 20 mV/s). Whereas the primary beam remains almost unaffected, the intensity of the diffracted beam can be gradually varied between the OFF and ON states. Movie S1 shows all frames for six cycles, including the corresponding voltammograms. The diffracted beam intensities that display hysteresis behavior, which enables powerless operation in the ON or OFF state, are depicted in fig. S7.

Our electrically switchable plasmonic nanoantennas and metasurfaces enabled by metallic polymers expand the functionality and performance of plasmonic-based electro-optical active devices and on-chip optical components. Fabrication from PEDOT:PSS is low cost and scalable, owing to the commercial availability of this polymer. We envision that our concept will be of importance in several distinct fields. Subwavelength-sized polymer nanoantennas will make a sizable impact in the development of displays and active optical components. One can address individual subwavelength pixels to push the pixel densities of emerging optical technologies to entirely new dimensions. Furthermore, in comparison with state-of-the-art metallic and dielectric nanoantennas, the polymer nanoantennas allow a previously unseen level of flexibility for fabrication of curved optical devices on flexible substrates. This functionality is necessary to achieve augmented and virtual reality technologies that work in transmission (e.g., on contact lenses or glasses). Ultimately, it may even be possible to achieve pixel densities >2000 lines/mm, which would support full-color holographic movies at a very large field of view. All of these advances will be aided by nanoantenna operation at only ±1 V, which is very favorable for compatibility with low-voltage complementary metal-oxide semiconductors (CMOS) chips (0 to 3.3 V) at moderate local electric fields. From a more fundamental standpoint, current efforts in quantum technology and methodology require intricate coupling and control schemes, which can be realized via, e.g., sophisticated switchable and reconfigurable metasurfaces, which are currently out of reach even with existing state-of-the-art techniques (*32*). Such structures would allow for an entirely new level of integration and miniaturization. From a basic research point of view, further studies into the working principle of the plethora of metallic and conductive polymers will allow for fine-tuning and manipulation of their properties. This includes shifting the plasma frequency, and thus the operation point, into the visible wavelength region; increasing the switching speed for cycling beyond the demonstrated video rate; and reducing degradation. Further detailed investigations on the influence of, e.g., oxygen and humidity during electrochemical switching, adjustments to the fabrication process, and electrochemical cell sealing in inert gas environments or the use of solid electrolytes should benefit the overall switching performance. Additionally, conductive polymers have recently proven to be stable for an extremely long time, with little to no degradation over $>10^7$ cycles at video-rate switching frequencies (*33*). Metallic polymers also offer the opportunity to gradually change the carrier density and hence the plasma frequency, which allows for gray-scale operation and thus opens another window of opportunity. These results, in combination with the shown hysteresis behavior that enables powerless operation in the ON or OFF state, suggest that extremely energy-efficient display devices could be realized in the future.

REFERENCES AND NOTES

1. I. Kim *et al*., *Nat. Nanotechnol.* **16**, 508–524 (2021).
2. A. H. Dorrah, N. A. Rubin, A. Zaidi, M. Tamagnone, F. Capasso, *Nat. Photonics* **15**, 287–296 (2021).
3. G. Y. Lee *et al*., *Nat. Commun.* **9**, 4562 (2018).
4. J. Park *et al*., *Nat. Nanotechnol.* **16**, 69–76 (2021).
5. W.-J. Joo *et al*., *Science* **370**, 459–463 (2020).
6. N. Yu *et al*., *Science* **334**, 333–337 (2011).
7. L. Huang *et al*., *Nano Lett.* **12**, 5750–5755 (2012).
8. A. V. Kildishev, A. Boltasseva, V. M. Shalaev, *Science* **339**, 1232009 (2013).
9. M. Khorasaninejad *et al*., *Science* **352**, 1190–1194 (2016).
10. S. Wang *et al*., *Nat. Nanotechnol.* **13**, 227–232 (2018).
11. A. Arbabi, Y. Horie, M. Bagheri, A. Faraon, *Nat. Nanotechnol.* **10**, 937–943 (2015).
12. X. Ni, N. K. Emani, A. V. Kildishev, A. Boltasseva, V. M. Shalaev, *Science* **335**, 427 (2012).
13. A. M. Shaltout, V. M. Shalaev, M. L. Brongersma, *Science* **364**, eaat3100 (2019).
14. M. Wuttig, H. Bhaskaran, T. Taubner, *Nat. Photonics* **11**, 465–476 (2017).
15. S. Chen *et al*., *Nat. Nanotechnol.* **15**, 35–40 (2020).
16. Q. Wang *et al*., *Nat. Photonics* **10**, 60–65 (2016).
17. J. Karst *et al*., *Sci. Adv.* **6**, eaaz0566 (2020).
18. F. Sterl *et al*., *Nano Lett.* **15**, 7949–7955 (2015).
19. Y. Wang *et al*., *Nat. Nanotechnol.* **16**, 667–672 (2021).
20. Y. Zhang *et al*., *Nat. Nanotechnol.* **16**, 661–666 (2021).

21. G. K. Shirmanesh, R. Sokhoyan, P. C. Wu, H. A. Atwater, *ACS Nano* **14**, 6912–6920 (2020).
22. J. Li, P. Yu, S. Zhang, N. Liu, *Nat. Commun.* **11**, 3574 (2020).
23. T. Xu *et al.*, *Nat. Commun.* **7**, 10479 (2016).
24. J. Peng *et al.*, *Sci. Adv.* **5**, eaaw2205 (2019).
25. S. Q. Li *et al.*, *Science* **364**, 1087–1090 (2019).
26. J. Ratzsch *et al.*, *J. Opt.* **22**, 124001 (2020).
27. Y. Leroux *et al.*, *ACS Nano* **2**, 728–732 (2008).
28. R. Kaissner *et al.*, *Sci. Adv.* **7**, eabd9450 (2021).
29. X. Yin *et al.*, *Light Sci. Appl.* **6**, e17016 (2017).
30. D. Neusser *et al.*, *Chem. Mater.* **32**, 6003–6013 (2020).
31. M. Wieland *et al.*, *Flex. Print. Electron.* **5**, 014016 (2020).
32. A. S. Solntsev, G. S. Agarwal, Y. S. Kivshar, *Nat. Photonics* **15**, 327–336 (2021).
33. K. Xiong *et al.*, *Adv. Mater.* **2103217**, e2103217 (2021).

ACKNOWLEDGMENTS

We thank F. Sterl and F. Mörz for contributions to the experimental setup and L.-S. Hornberger and B. Gompf for contributions to the ellipsometry measurements. **Funding:** We acknowledge financial support from Baden-Wuerttemberg-Stiftung (Opterial), the European Research Council (ERC Advanced Grant Complexplas and ERC PoC Grant 3DPrintedOptics), Bundesministerium für Bildung und Forschung, Deutsche Forschungsgemeinschaft (SPP1839 Tailored Disorder and GRK2642 Photonic Quantum Engineers), the Institute of Quantum Science and Technology (IQST), the University of Stuttgart (Terra Incognita), the Carl Zeiss Foundation, and the Vector Foundation. **Author contributions:** J.K., M.H., and H.G. conceived the project. J.K. and M.U. fabricated the samples. J.K. and M.F. performed the measurements. J.K. carried out the data analysis. C.D., C.M., and T.S. contributed to the experimental setup, sample fabrication, and measurements. H.G., M.H., and S.L. supervised the project. All authors discussed the results and contributed to manuscript writing. **Competing interests:** J.K., M.H., and H.G. have filed an international patent related to the topic of this work (PCT/EP2021/069202). **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj3433
Materials and Methods
Figs. S1 to S7
Reference (*34*)
Movie S1

10 July 2021; accepted 13 September 2021
10.1126/science.abj3433

MAGNETISM

Intrinsic 2D-XY ferromagnetism in a van der Waals monolayer

Amilcar Bedoya-Pinto[1]*†, Jing-Rong Ji[1]†, Avanindra K. Pandeya[1], Pierluigi Gargiani[2], Manuel Valvidares[2], Paolo Sessi[1], James M. Taylor[3], Florin Radu[3], Kai Chang[1]*, Stuart S. P. Parkin[1]*

The physics and universality scaling of phase transitions in low-dimensional systems has historically been a topic of great interest. Recently, two-dimensional (2D) materials exhibiting intriguing long-range magnetic order have been in the spotlight. Although an out-of-plane anisotropy has been shown to stabilize 2D magnetic order, the demonstration of a 2D magnet with in-plane rotational symmetry has remained elusive. We constructed a nearly ideal easy-plane system, a single $CrCl_3$ monolayer on graphene/6H-SiC(0001), and observed robust ferromagnetic ordering with critical scaling characteristic of a 2D-XY system. These observations indicate the realization of a finite-size Berezinskii-Kosterlitz-Thouless phase transition in a large-area, quasi–free-standing van der Waals monolayer magnet with an XY universality class. This offers a material platform to host 2D superfluid spin transport and topological magnetic textures.

How reduced dimensionality affects physical properties is a question that has fascinated researchers for decades. Recent technological advances have enabled the study of truly one- and two-dimensional systems. In this context, theoretical models of low-dimensional magnetism have been revisited. It was recognized early that the absence of long-range magnetic order in a low-dimensional system ($d < 3$) with continuous symmetry, postulated by Mermin and Wagner (*1*, *2*), could be remedied by the existence of a sizable magnetic anisotropy. In 2D systems with easy-plane anisotropy, an unusual long-range order is expected to emerge as a result of the formation of magnetic vortex-antivortex or chirally opposing domain wall–bound pairs, theoretically described by the Berezinskii-Kosterlitz-Thouless (BKT) formalism (*3*–*5*). This implies the occurrence of a phase transition even in a 2D magnetic system with continuous rotational symmetry $O(2)$. This theory was later adapted to account for symmetry-breaking fields (*6*–*8*), finite size effects (*9*, *10*), and spin-wave interactions with magnetic vortices (*11*). The work in (*6*), (*7*), (*9*), and (*11*) allowed the quantification of the order parameters and critical scaling near the phase transition for each scenario. A renormalization group approach was used to show that the magnetization of finite-sized 2D easy-plane systems—also called 2D-XY—scales with a critical exponent of $\beta = 3\pi^2/128 \approx 0.231$, defining a fingerprint of the XY universality class (*9*, *12*). Perturbations to the rotational invariance of the spins in the plane—such as the symmetry and strength of magnetocrystalline fields—have been shown to strongly influence the critical exponents and the universality class. Whereas a uniaxial in-plane anisotropy (XYh_2) drives the system into an Ising-type behavior ($\beta = 0.125$), the effect of a four-fold symmetry (XYh_4) is highly dependent on the crystal field strength, with critical exponents ranging between Ising and XY universality classes (*13*). In contrast, the hexagonal symmetry (XYh_6) barely affects the magnetic behavior (*4*, *6*, *7*) even in the limit of strong crystal fields (*13*), such that the system remains in the XY class, thereby enabling the observation of a BKT-type phase transition.

A large number of experiments on quasi-2D magnetic systems, such as monolayers on crystalline metallic surfaces or bulk layered magnets with small interplanar exchange interactions, have contributed to the assessment of the aforementioned theories [see (*13*, *14*) for reviews]. However, in those cases, the substrate interaction through bonding and hybridization (monolayers on surfaces) and the small amount of interplanar exchange (layered magnets) were conditions that precluded the realization of an ideal 2D system. Recent advances in the isolation and preparation of crystalline monolayers from innately layered magnets, via exfoliation (*15*–*17*) or molecular beam epitaxy (MBE) (*18*), allow for more detailed investigations of low-dimensional critical phenomena. So far, ferromagnetic ordering has been demonstrated in CrI_3, $CrBr_3$, and Fe_3GeTe_2 monolayers (*15*, *16*, *18*), all of them stabilized by a substantial out-of-plane uniaxial anisotropy that places them close to the Ising universality class, as discussed in recent reviews (*19*–*22*). A careful analysis of some of these systems, such as ferromagnetic $CrBr_3$ (*23*) or antiferromagnetic $NiPS_3$ (*24*), rather suggests XXZ-type magnetic behavior, the latter losing magnetic ordering in the monolayer regime. $CrCl_3$, on the other hand, is an in-plane antiferromagnet in the bulk form (*25*, *26*), resulting from an alternation between the magnetic moments in each individual $CrCl_3$ layer, which are ferromagnetically aligned in-plane. This leads to the question of whether a single-layer $CrCl_3$ would order ferromagnetically—or order magnetically at all—given its weak anisotropy and the lack of interplanar exchange in the monolayer limit.

Recent reports on the magnetic properties of $CrCl_3$ exfoliated flakes reach down to the bilayer regime (*27*–*30*), where the antiferromagnetic interlayer exchange is still preserved, whereas the determination of the magnetic properties of a monolayer has remained elusive. One reason is that the magnetic characterization of the monolayer flakes relies on indirect methods, which require device fabrication for tunneling magnetoresistance (*27*–*30*) or Hall-micromagnetometry experiments (*23*). Direct methods such as the magneto-optical

[1]NISE Department, Max Planck Institute of Microstructure Physics, Halle, Germany. [2]ALBA Synchrotron Light Source, Barcelona, Spain. [3]Helmholtz-Zentrum für Materialien und Energie, Berlin, Germany.
***Corresponding author. Email: abedoya@mpi-halle.mpg.de (A.B.-P.); changkai@baqis.ac.cn (K.C.); stuart.parkin@mpi-halle.mpg.de (S.S.P.P.)**
†These authors contributed equally to this work.

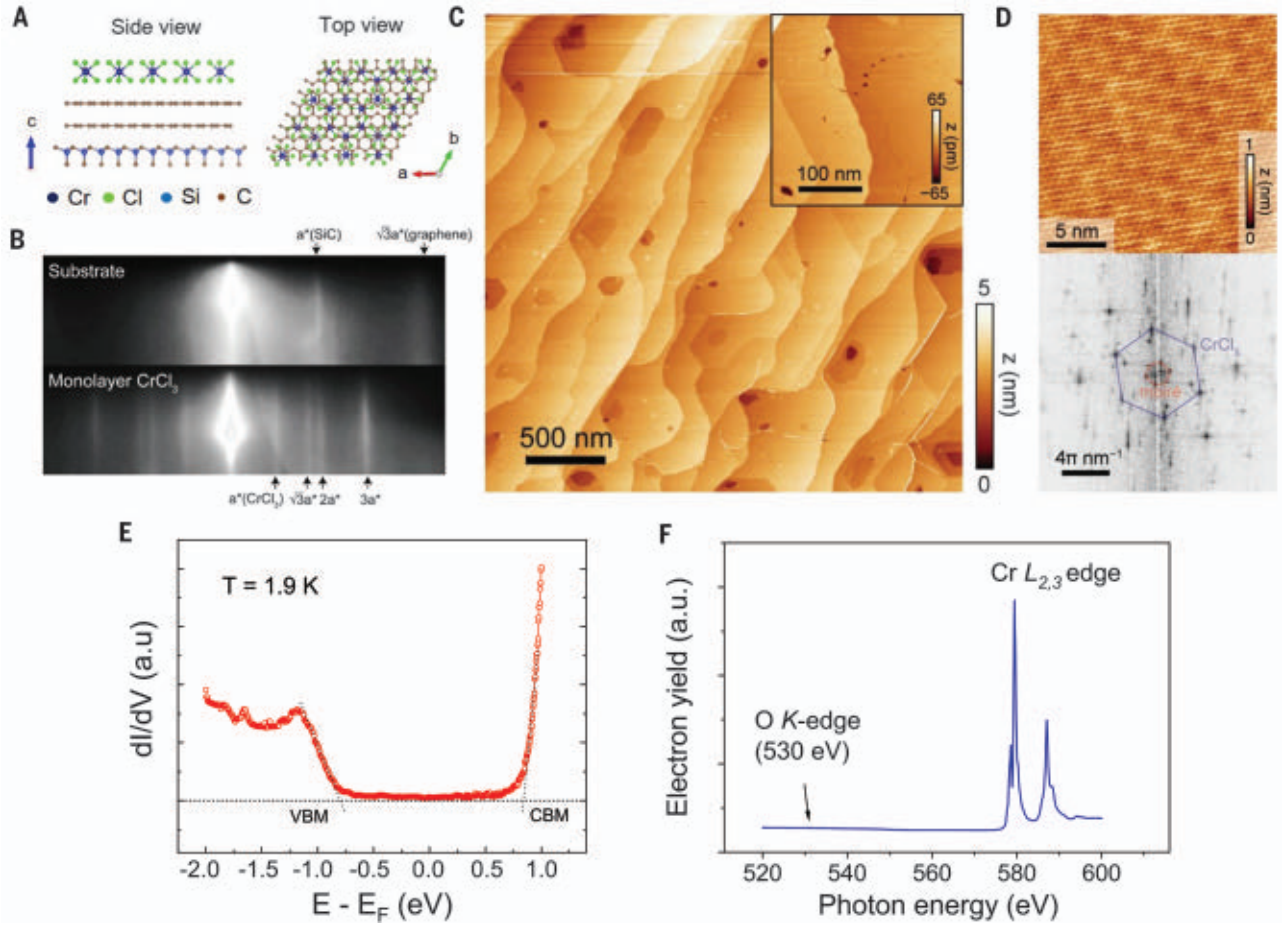


Fig. 1. Structural and electronic properties of a $CrCl_3$ monolayer. (**A**) Schematic crystal structure of $CrCl_3$/graphene/6H-SiC layers in cross section and top-view configurations. (**B**) In situ RHEED pattern of the substrate and monolayer $CrCl_3$ grown by MBE, along Γ-*M* of graphene (Γ-*K* of SiC). Streaks from two different high-symmetry directions of $CrCl_3$ imply a twisted in-plane orientation of the grains. (**C**) STM topography of monolayer $CrCl_3$ grown on graphene/6H-SiC(0001), indicating a homogeneous coverage on long length scales. Set points: Sample bias voltage V = +1.2 V, tunneling current I = 5 pA. Inset: A magnified topography image, which reveals the grain boundaries. (**D**) Atom-resolved image of the $CrCl_3$ lattice featuring a moiré pattern (top) and its Fourier-transformed image (bottom). Set points: V = +0.1 V, I = 100 pA. The moiré pattern corresponds to a 23.8° rotation between the hexagonal unit cell of $CrCl_3$ and graphene. Images in (C) and (D) were acquired at room temperature. (**E**) *dI/dV* spectrum at the surface of monolayer $CrCl_3$, taken at 1.9 K. The estimated bandgap is 1.6 eV, obtained by linearly extrapolating the sharp increase in signal at positive and negative energies to intersect the energy axis; a.u., arbitrary units. VBM, valence band maximum; CBM, conduction band minimum. (**F**) XAS near the O *K* and Cr $L_{2,3}$ edge region, ruling out the presence of oxygen in the surface and highlighting a sharp Cr^{3+} absorption white line.

Kerr effect suffer from small signals due to sample dimensions and unfavorable magneto-optical coefficients in $CrCl_3$ (*31*); volumetric magnetometry methods are not suitable for exploring monolayer samples owing to large background signals. Here, we circumvented these difficulties by preparing a large-area, homogeneous $CrCl_3$ monolayer on graphene/6H-SiC(0001) by MBE, and by measuring its intrinsic magnetic properties in situ by element-specific x-ray magnetic circular dichroism (XMCD).

The structure of the $CrCl_3$ monolayer grown on graphitized 6H-SiC(0001) by MBE features a van der Waals gap to the substrate and a crystalline layer with an in-plane hexagonal lattice (Fig. 1A). The Cr atom is coordinated in an octahedral configuration to the neighboring Cl atoms (i.e., Cr-Cl bonds are off-plane, and the Cr atoms form a honeycomb lattice). In situ reflection high-energy electron diffraction (RHEED) patterns show that the $CrCl_3$ films have a surface perpendicular to the *c* axis and that the film microstructure presents in-plane twisted domains, as seen in the multiple diffraction streaks corresponding to a^* and $\sqrt{3}a^*$ $CrCl_3$ lattice periodicities along a high-symmetry direction (e.g., Γ-*M* of graphene; Fig. 1B). This means that a strict six-fold symmetry is no longer guaranteed on long-range length scales. Figure 1C shows a 3-μm scanning tunneling microscopy (STM) topographic image of a $CrCl_3$ monolayer, indicating a homogeneous coverage of the graphene substrate, and a crystal grain size of 200 to 300 nm (Fig. 1C, inset). Inside a $CrCl_3$ grain, the atom-resolved STM image (Fig. 1D) displays the $CrCl_3$ hexagonal lattice superimposed on a clear moiré pattern, which corresponds to a twist angle of 23.8° (fig. S1) between the $CrCl_3$ monolayer and the graphene substrate. Various moiré patterns were observed in different grains, which further supports random in-plane grain orientations (fig. S2) consistent with a weak interaction with the graphene substrate. The local electronic properties of the $CrCl_3$ monolayer, mapped by scanning tunneling spectroscopy (Fig. 1E), reveal a bandgap of ~1.6 eV, which is close to what is predicted from ab initio calculations (*32*) but substantially lower than the bandgap (3.0 eV) measured in bulk samples (*33*). The Fermi energy lies in the middle of the gap, indicating that the electronic properties of the MBE-grown monolayer are intrinsic, with a low concentration of defects/dopants and a negligible charge transfer effect from the substrate. Prior to the in situ investigation of the magnetic properties of the samples, we performed x-ray absorption spectroscopy (XAS) measurements to assess

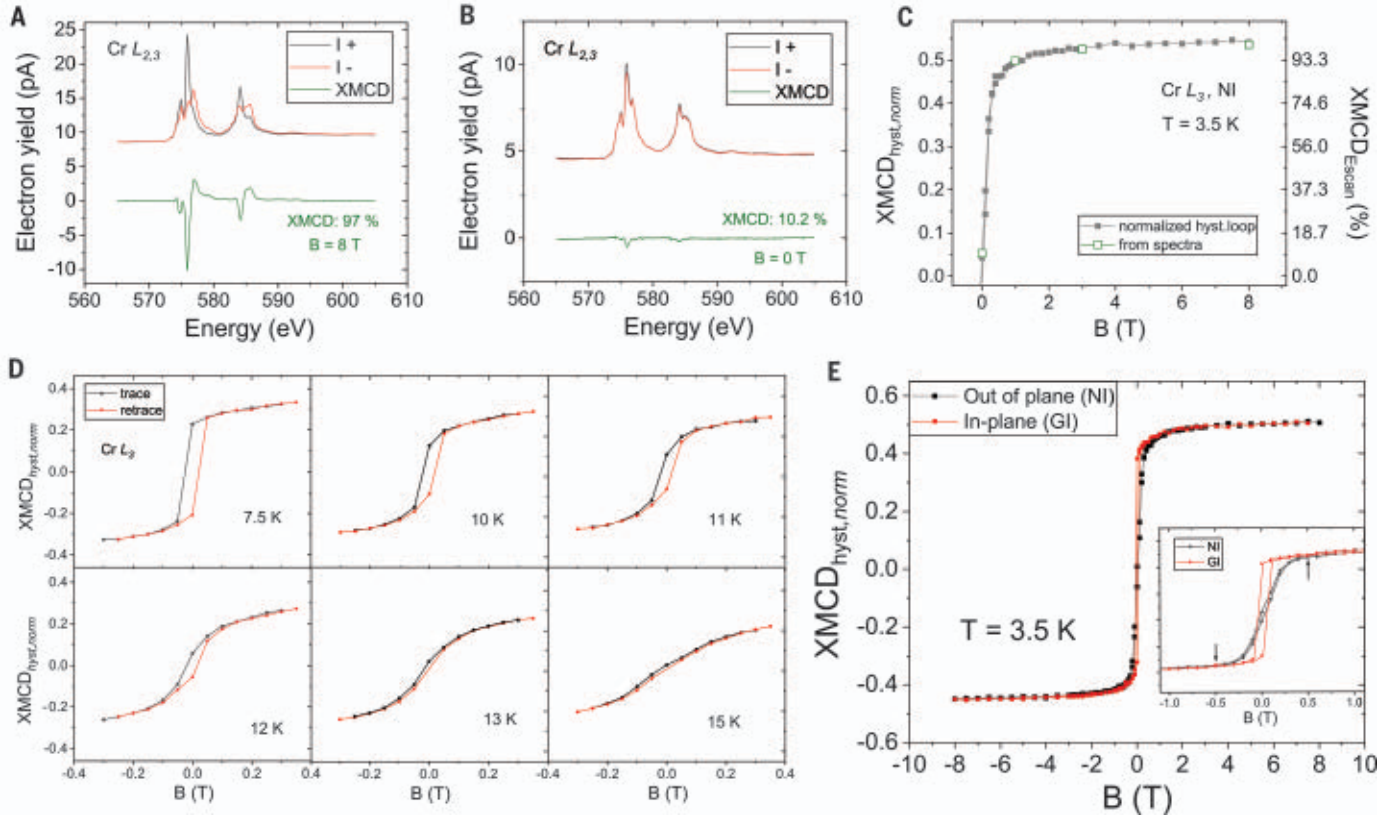


Fig. 2. Element-specific magnetic properties of $CrCl_3$ monolayer measured by x-ray magnetic dichroism. (**A** and **B**) X-ray absorption spectra with different photon helicities and the resulting difference (XMCD) for the Cr $L_{2,3}$ edge at B = 8 T (A) and at zero magnetic field (B), applied out-of-plane. A very large dichroic signal (close to 100%) is observed. (**C**) Comparison between the XMCD magnitudes extracted from the energy spectra and from XMCD hysteresis loop sequences, showing consistency between the two methods; NI, normal incidence. (**D**) XMCD signal as a function of in-plane magnetic field, taken at various temperatures. Remanence and coercive field become negligible at 13 K. (**E**) XMCD hysteresis loops taken in normal and grazing (GI) incidence. Inset: A snap of the low-field region reveals the anisotropy fields (0.5 T).

the chemical integrity of the $CrCl_3$ surface after transfer via an ultrahigh-vacuum taxi chamber to the synchrotron beamline (*34*). We did not observe any trace of oxygen contamination but rather a well-defined Cr x-ray absorption line, with sharp multiplet peaks characteristic of a Cr^{3+} ion in an octahedral configuration (Fig. 1F).

Figure 2 summarizes the key magnetic features of the $CrCl_3$ monolayer, revealed from XMCD measurements: (i) large XMCD signals at the Cr $L_{2,3}$ edge (close to 100% at L_2 at saturation fields) indicative of sizable magnetic moments on the Cr^{3+} ions (as well as at the Cl sites; see fig. S3), (ii) a field-dependent magnetization with nonzero remanence and coercive fields typical of ferromagnetic ordering, (iii) a detectable magnetic anisotropy favoring an in-plane easy axis, and (iv) a ferromagnetic-to-paramagnetic transition around 13 K. As shown in Fig. 2A, the x-ray absorption spectra at the Cr $L_{2,3}$ edge taken with different photon helicities (right- and left-handed, I+ and I– for simplicity) show a huge difference, yielding values of nearly 100% at high magnetic fields (8 T) and normal-incidence geometry (field perpendicular to the plane). Interestingly, the XMCD signal is still detectable in zero magnetic field (about 10%), which indicates the presence of a remanent magnetic moment at low temperatures (3.5 K) (Fig. 2B). This high XMCD contrast was crucial to performing measurements of the Cr L_3 edge at different magnetic fields (hysteresis loops) with a high level of accuracy. Figure 2C shows the correspondence of the XMCD intensity extracted from the energy spectra (in percent, normalized to the XAS L_3 edge jump) to that acquired during a typical hysteresis loop measurement [see (*34*) for details of the XMCD data acquisition and the relation to the total magnetic moment]. Using sum-rule analysis (fig. S4), we extracted a total magnetic moment of 2.9 μB/Cr, very close to the theoretically expected value of 3 μB for a trivalent Cr valence ($3d^3$). Although we also found a remarkably high XMCD signal (30%) at the chlorine edge (fig. S3), a reliable determination of the Cl moments via sum-rule analysis was not possible because of the small energy separation of L_3 and L_2 lines resulting from the weak spin-orbit coupling of Cl. Assuming a ferromagnetic superexchange coupling mechanism (90° hopping path) between Cr spins and Cl ligands in $CrCl_3$, as expected from the Goodenough-Kanamori rules (*35*, *36*), it is very likely that the spin moment of Cl aligns antiparallel to that of Cr. Indeed, the small reduction of the Cr^{3+} magnetic moment (2.9 μB instead of 3 μB) might arise from a certain degree of p-d admixture of Cr and Cl orbitals in the superexchange path. The resulting exchange coupling is small (J_{ex} ~ 0.67 to 0.75 meV by mean-field approximation) inferred from the ordering temperature observed in the monolayer samples (see table S1). Figure 2D displays the temperature evolution of the Cr^{3+} XMCD hysteresis loops under an in-plane applied magnetic field (grazing-incidence geometry), showing that the squareness of the hysteresis progressively diminishes as the temperature is raised; at 13 K, both remanence and coercive fields vanish, indicative of a ferro- to paramagnetic phase transition.

To evaluate the magnitude of the underlying magnetic anisotropy, we took field-dependent XMCD hysteresis loops at normal- and grazing-incidence angles at 3.5 K (Fig. 2E), evidencing a clear in-plane magnetic easy axis. The extracted anisotropy fields (Fig. 2E, inset) are on the order of 0.5 to 0.6 T (corresponding to anisotropy energies of 0.09 to 0.11 meV). These values are much larger than those anticipated from

first-principles calculations of a $CrCl_3$ monolayer (*32*, *37*, *38*), which range from 0.031 to 0.055 meV/Cr, and also turn out to be slightly larger than the critical field anisotropy in the antiferromagnetic bulk (0.3 to 0.5 T) (*26*, *33*) and in bilayer $CrCl_3$ (0.23 T) (*29*). A stable ferromagnetic ordering was not expected a priori when reducing the dimensionality of the few-layer antiferromagnetic system down to the monolayer, as it was unclear whether the removal of the interlayer exchange interaction would leave the intralayer ferromagnetic behavior intact. In fact, the magnitudes of the calculated magnetic anisotropy energies for a single $CrCl_3$ monolayer were so low that some calculations (*32*, *37*) even predict a marginally favorable perpendicular magnetic anisotropy instead of the expected in-plane one. Our observation of a clear in-plane easy axis in monolayer $CrCl_3$ provides a decisive answer to this puzzle and consolidates its classification as a two-dimensional in-plane ferromagnet. The enhanced magnetic anisotropy (0.09 to 0.11 meV) cannot be simply understood by a dominant shape anisotropy [upper bound of 0.052 meV calculated in (*39*)]—as is the case in few-layer exfoliated $CrCl_3$ samples—but rather is explained by sizable contributions of the single-ion anisotropy (ΔE_{SI}) and magnetic exchange terms (ΔE_{ME}) (*40*, *41*). In this regard, we found a detectable orbital-moment anisotropy (ΔL = 0.04) using sum-rule evaluation of the XMCD spectra at grazing and normal incidence (fig. S4) as well as a substantial deviation of the Cr-Cl-Cr bond plane angle ($\delta\theta = 10°$) via angle-dependent x-ray absorption measurements (fig. S5), the latter related to a trigonal distortion of the ideal $CrCl_3$ octahedral environment. Both factors have a direct impact on the magnitude of the single-ion and magnetic exchange anisotropy in $CrCl_3$ and thus lie at the origin of the enhanced anisotropy [see (*34*) for an extended discussion and qualitative estimation of the relevant magnetic anisotropy terms]. We expect that this substantial trigonal distortion and the related orbital moment anisotropy are induced during the nucleation and epitaxial growth process of our monolayer samples. We note that our few-layer $CrCl_3$ samples grown by MBE show magnetic behavior similar to that reported in exfoliated flakes (*27*–*30*), exhibiting signatures of spin-flop transitions at low fields (*30*) and giant interlayer exchange enhancement (*28*, *30*) in the form of large critical fields (fig. S6).

To characterize the magnetic behavior below and above the Curie temperature T_c and to gain insight into the nature of the phase transition, we acquired the low-field dependence of the Cr L_3 edge XMCD between 12 K and 15 K in the easy-axis direction (see fig. S7 for the detailed datasets close to the phase transition). Concomitant with the disappearance of the hysteresis at 13 K (Fig. 2D), the M(H) curves evolve into a softer S-shape above T_c and

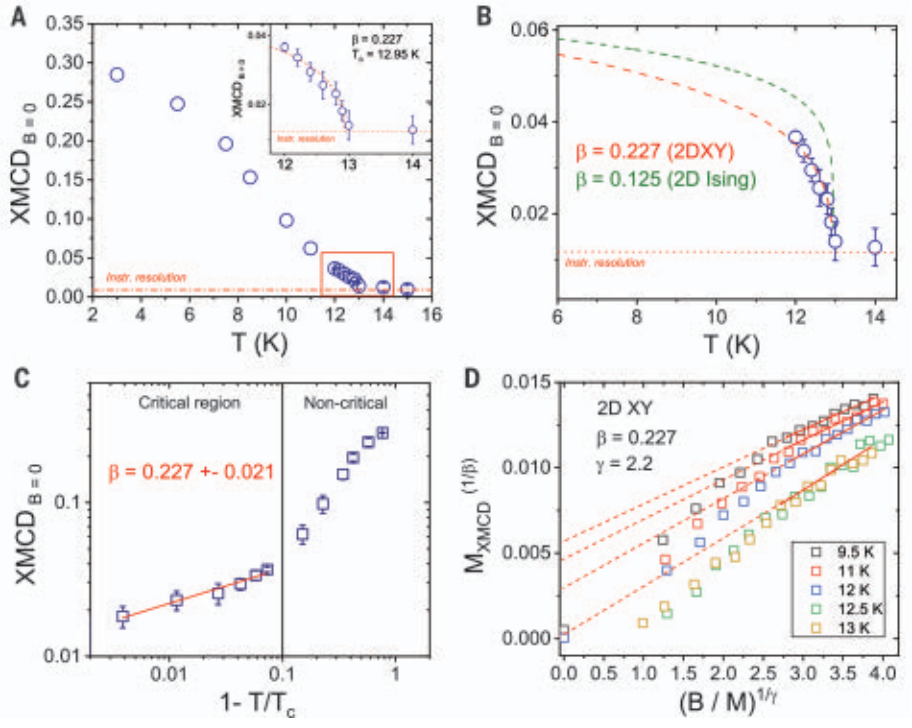


Fig. 3. Scaling behavior and critical exponents. (**A**) XMCD values at zero magnetic field (remanence) as a function of temperature, extracted from the hysteresis loops. Inset: The critical scaling fit close to the phase transition determines the exponent β = 0.227. (**B**) The fits using the 2D Ising (green) and 2D-XY models (red) are shown; the latter is consistent with our data. (**C**) Logarithmic representation of XMCD magnitude versus reduced temperature [1 – (T/T_c)], for an optimal visualization of the accuracy of the inferred critical exponent β and the definition of the critical temperature region. (**D**) Modified Arrott plots of the field-dependent XMCD data at various temperatures (β = 0.227, γ = 2.2), matching the 2D-XY model predictions. M is the magnetic moment, proportional to the XMCD intensity. Extrapolation of the linear fits at high fields is drawn as a guide to the eye to visualize the y-intercept evolution toward the phase transition (y = 0 at T_c).

move toward a linear dependence, entering the paramagnetic regime. The magnetization is a good order parameter and scales as $M = M_0[1 - (T/T_c)]^{\beta}$, where β is the critical exponent that determines the magnetic universality class. In two-dimensional systems, the isotropic Heisenberg system (n = 3) does not have any spontaneous order, and the Ising (n = 1) and XY (n = 2) models are the relevant scenarios for $CrCl_3$, where n is the dimensionality of the spin degree of freedom. Although often understood in terms of an out-of-plane easy axis, the Ising universality can also be found in in-plane magnetized systems (*42*, *43*) as long as there is only one preferred (uniaxial) magnetization direction. To perform the analysis of the critical exponent β, we rely on the evolution of the remanent magnetic moment m_r [proportional to the XMCD L_3 intensity at zero field (*34*)] as a function of temperature (Fig. 3A). The inset shows the fitting of the data close to the phase transition (12 to 13 K), yielding values of β = 0.227 ± 0.021 and T_c = 12.95 ± 0.03 K, matching well with the expected value (β = 0.231) of the 2D-XY model (*9*). If the exponent β is set to equal 0.125, corresponding to Ising-type universality (Fig. 3B, green line), the agreement with the experimental data is rather poor. A better visualization of the temperature region of critical behavior is given by a log-log representation of the magnetization versus reduced temperature [1 – (T/T_c)], as shown in Fig. 3C. The region where the critical scaling holds occurs in a range of reduced temperatures between 0.1 and 0.003, a regime typically used in critical phenomena studies [see (*14*) for a compendium].

To further assess the 2D-XY behavior in our samples, we carried out a second analysis procedure, the so-called Arrott-Noakes plots (*44*), from the field- and temperature-dependent XMCD data. In this representation, the magnetization and susceptibility power-law scaling are visualized together across the phase transition, and a consistent set of the critical exponents β and γ, corresponding to the magnetization and susceptibility, can be deduced. This approach is widely used to distinguish between the various magnetic interaction models (e.g., Heisenberg, Ising, XY, Potts) in three and two dimensions, each of which have a specific set of critical exponents. The determination of the critical exponent γ by an Arrott-plot analysis of the temperature-dependent XMCD data is shown in Fig. 3D. The analysis of the Arrott plots is consistent with the previously inferred value

of β (0.227) and also defines the susceptibility critical exponent as γ = 2.2, the latter falling reasonably in the range estimated from the high-temperature series expansions for the 2D-XY model, γ = 2.4 ± 0.3 (*45*). However, although the linear fits and the evolution of the intercepts are reasonable for this set of critical exponents, the Arrott-plot analysis is not as accurate as the temperature-dependent magnetization to determine T_c (i.e., the high-field data of 12.5 K and 13 K fall under the same fitting range). [See (*34*) for more details about issues in the determination of the susceptibility exponent γ of 2D-XY systems]. These results have been reproduced on a second sample with a slightly different T_c but showing the same XY universality scaling, demonstrating that monolayer $CrCl_3$ grown by MBE is an easy-plane magnet (see table S1 and fig. S8 for the additional datasets).

The very good agreement between the critical exponent β and the theoretically expected value is attributed to the 2D-XY features of our samples: vanishing in-plane crystalline anisotropy fields, negligible substrate interaction, and a relatively large magnetic coherence size L of the $CrCl_3$ monolayer owing to a homogeneous, atomically flat morphology on the graphene substrate. (Other relevant 2D universality scaling models fail to describe our data; see fig. S9.) The observed 2D-XY ordering in monolayer $CrCl_3$ suggests that the nature of the phase transition is of the BKT type. This universal phase transition has been observed in other physical systems such as superconductors (*46*, *47*), atomic surface reconstructions (*48*), and quasi-condensates of atomic gases (*49*). In magnetic systems, this transition has been argued to occur in metallic monolayers grown on crystalline substrates (*50*, *51*) and in layered bulk magnets (*52*). However, the strong substrate interaction in the first case, and interlayer exchange in the second case, might cause a departure from ideal 2D-XY behavior where the BKT transition is expected. We believe that our $CrCl_3$ system, a quasi–free-standing van der Waals monolayer, is a more suitable setting for a magnetic BKT transition to take place. Further studies on finite-size effects and exponential scaling of the magnetic correlation length are needed to fully assess this conjecture. In a very recent theoretical work, a transport signature of the BKT transition in magnetic easy-plane systems has been formulated (*53*), which perfectly applies to our $CrCl_3$ monolayer and will yield additional insights on the spin transport phenomena of 2D easy-plane magnets, which are expected to reach the spin superfluidity regime (*54*).

We have shown that a $CrCl_3$ monolayer, featuring van der Waals coupling to the graphene substrate and hexagonal symmetry in the plane but without strict single-crystalline long-range order, constitutes a nearly ideal realization of a 2D-XY magnetic system. The van der Waals nature at the $CrCl_3$/substrate interface minimizes effects such as hybridization, bonding, and substrate-driven crystalline anisotropy fields; these factors can drive the system into an Ising or anisotropic Heisenberg type, or may even cause a departure from 2D behavior by strong substrate interaction. The in-plane hexagonal crystal symmetry (XYh6) of $CrCl_3$, on the other hand, is the least perturbative for the XY model (*6*) and is additionally diminished in our experimental system as a result of the in-plane twisting of crystalline domains. As compared with bulk layered magnets, our monolayer system has no interplanar exchange coupling $J_{ex,inter}$, which is also a perturbation of ideal 2D-XY behavior and dimensionality. A follow-up study of possible interest would consist of gently turning on in-plane anisotropy fields by choosing different rigid substrates with strong crystalline fields or a given step terrace morphology, and seeing how this affects the 2D-XY behavior of the monolayer. In that way, recipes to attain a crossover to Ising-type (uniaxial anisotropy) behavior may be developed, offering a useful pathway to the design of on-demand magnetic anisotropies for functional spintronic devices based on 2D materials. On a more fundamental side, the finite-size effects of the BKT phase transition can be studied by modifying the grain size and percolation behavior of the MBE-grown monolayers. Moreover, an epitaxy-controlled trigonal distortion could be used as an experimental parameter to set the balance between single-ion anisotropy and dipolar interaction in $CrCl_3$, which is crucial for stabilizing topological spin textures such as merons (*55*, *56*) in two dimensions.

REFERENCES AND NOTES

1. P. C. Hohenberg, *Phys. Rev.* **158**, 383–386 (1967).
2. N. D. Mermin, H. Wagner, *Phys. Rev. Lett.* **17**, 1133–1136 (1966).
3. V. Berezinskiĭ, *Sov. Phys. JETP* **34**, 610 (1972).
4. J. M. Kosterlitz, D. J. Thouless, *J. Phys. C* **6**, 1181–1203 (1973).
5. J. M. Kosterlitz, *J. Phys. C* **7**, 1046–1060 (1974).
6. J. V. José, L. P. Kadanoff, S. Kirkpatrick, D. R. Nelson, *Phys. Rev. B* **16**, 1217–1241 (1977).
7. K. Y. Szeto, G. Dresselhaus, *Phys. Rev. B* **32**, 3186–3193 (1985).
8. S. T. Bramwell, P. C. W. Holdsworth, J. Rothman, *Mod. Phys. Lett. B* **11**, 139–148 (1997).
9. S. T. Bramwell, P. C. W. Holdsworth, *J. Phys. Condens. Matter* **5**, L53–L59 (1993).
10. K. Y. Szeto, S. T. Chen, G. Dresselhaus, *Phys. Rev. B* **32**, 4628–4638 (1985).
11. J. Villain, *J. Phys.* **36**, 581–590 (1975).
12. S. T. Bramwell, P. C. W. Holdsworth, *Phys. Rev. B* **49**, 8811–8814 (1994).
13. A. Taroni, S. T. Bramwell, P. C. W. Holdsworth, *J. Phys. Condens. Matter* **20**, 275233 (2008).
14. C. A. F. Vaz, J. A. C. Bland, G. Lauhoff, *Rep. Prog. Phys.* **71**, 056501 (2008).
15. B. Huang *et al.*, *Nature* **546**, 270–273 (2017).
16. Z. Fei *et al.*, *Nat. Mater.* **17**, 778–782 (2018).
17. C. Gong *et al.*, *Nature* **546**, 265–269 (2017).
18. W. Chen *et al.*, *Science* **366**, 983–987 (2019).
19. K. S. Burch, D. Mandrus, J. G. Park, *Nature* **563**, 47–52 (2018).
20. C. Gong, X. Zhang, *Science* **363**, eaav4450 (2019).
21. M. Gibertini, M. Koperski, A. F. Morpurgo, K. S. Novoselov, *zNat. Nanotechnol.* **14**, 408–419 (2019).
22. B. Huang *et al.*, *Nat. Mater.* **19**, 1276–1289 (2020).
23. M. Kim *et al.*, *Nat. Electron.* **2**, 457–463 (2019).
24. K. Kim *et al.*, *Nat. Commun.* **10**, 345 (2019).
25. B. Kuhlow, *Phys. Status Solidi* **72**, 161–168 (1982).
26. M. A. McGuire *et al.*, *Phys. Rev. Mater.* **1**, 014001 (2017).
27. X. Cai *et al.*, *Nano Lett.* **19**, 3993–3998 (2019).
28. D. R. Klein *et al.*, *Nat. Phys.* **15**, 1255–1260 (2019).
29. H. H. Kim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11131–11136 (2019).
30. Z. Wang *et al.*, *Nat. Nanotechnol.* **14**, 1116–1122 (2019).
31. A. Molina-Sánchez, G. Catarina, D. Sangalli, J. Fernandez-Rossier, *J. Mater. Chem. C* **8**, 8856–8863 (2020).
32. L. Webster, J. A. Yan, *Phys. Rev. B* **98**, 144411 (2018).
33. M. Abramchuk *et al.*, *Adv. Mater.* **30**, e1801325 (2018).
34. See supplementary materials.
35. J. B. Goodenough, *J. Phys. Chem. Solids* **6**, 287–297 (1958).
36. J. Kanamori, *J. Phys. Chem. Solids* **10**, 87–98 (1959).
37. W. B. Zhang, Q. Qu, P. Zhu, C. H. Lam, *J. Mater. Chem. C* **3**, 12457–12468 (2015).
38. B. Yang, X. Zhang, H. Yang, X. Han, Y. Yan, *J. Phys. Chem. C* **123**, 691–697 (2019).
39. F. Xue, Y. Hou, Z. Wang, R. Wu, *Phys. Rev. B* **100**, 224429 (2019).
40. D. H. Kim *et al.*, *Phys. Rev. Lett.* **122**, 207201 (2019).
41. J. L. Lado, J. Fernández-Rossier, *2D Mater.* **4**, 035002 (2017).
42. C. H. Back *et al.*, *Nature* **378**, 597–600 (1995).
43. Z. Q. Qiu, J. Pearson, S. D. Bader, *Phys. Rev. Lett.* **67**, 1646–1649 (1991).
44. A. Arrott, J. E. Noakes, *Phys. Rev. Lett.* **19**, 786–789 (1967).
45. J. Rogiers, E. W. Grundke, D. D. Betts, *Can. J. Phys.* **57**, 1719–1730 (1979).
46. D. J. Resnick, J. C. Garland, J. T. Boyd, S. Shoemaker, R. S. Newrock, *Phys. Rev. Lett.* **47**, 1542–1545 (1981).
47. M. Mondal *et al.*, *Phys. Rev. Lett.* **107**, 217003 (2011).
48. D. H. Baek, J. W. Chung, W. K. Han, *Phys. Rev. B* **47**, 8461–8464 (1993).
49. Z. Hadzibabic, P. Krüger, M. Cheneau, B. Battelier, J. Dalibard, *Nature* **441**, 1118–1121 (2006).
50. H. J. Elmers, J. Hauschild, G. H. Liu, U. Gradmann, *J. Appl. Phys.* **79**, 4984 (1996).
51. J. Atchison, A. Bhullar, B. Norman, D. Venus, *Phys. Rev. B* **99**, 125425 (2019).
52. J. Als-Nielsen, S. T. Bramwell, M. T. Hutchings, G. J. McIntyre, D. Visser, *J. Phys. Condens. Matter* **5**, 7871–7892 (1993).
53. S. K. Kim, S. B. Chung, *SciPost Phys.* **10**, 068 (2021).
54. S. Takei, Y. Tserkovnyak, *Phys. Rev. Lett.* **112**, 227201 (2014).
55. X. Lu, R. Fei, L. Zhu, L. Yang, *Nat. Commun.* **11**, 4724 (2020).
56. M. Augustin, S. Jenkins, R. F. L. Evans, K. S. Novoselov, E. J. G. Santos, *Nat. Commun.* **12**, 185 (2021).
57. Replication data for this report can be found in Zenodo. DOI: 10.5281/zenodo.5207282 (2021).

ACKNOWLEDGMENTS

A.B.-P. thanks the HZB and CELLS-ALBA for the allocation of synchrotron radiation beamtime under proposals 192-08773-ST and 202-09917-ST (HZB) and 2019-093862 (CELLS-ALBA). We thank F. Küster for assistance with the LT-STM measurements and C. Luo, K. Chen, S. Thakur, and S. Rudorff for technical support at BESSY. **Funding:** F.R. and A.B.-P. acknowledge financial support for the VEKMAG project and for the PM2-VEKMAG beamline by the German Federal Ministry for Education and Research (BMBF 05K10PC2, 05K10WR1, 05K10KE1) and by HZB. M.V. and P.G. acknowledge additional beamtime through ALBA IHR and proposal ID 2019023487-2, and funding via grants FIS2016-78591- C3-2-R (AEI/FEDER, UE) and FlagEra Sograph MEM PCI2019- 111908-2. K.C. was funded by National Natural Science Foundation of China (grant 12074038). **Author contributions:** A.B.-P., K.C., and S.S.P.P. conceived of the study, and A.B.-P. was the lead researcher. A.B.-P. carried out the magnetic characterization by XAS/XMCD, analyzed the data, and wrote the manuscript. J.-R.J. grew the samples for the beamtimes and performed the in situ RHEED and STM characterization. K.C. initiated and optimized the substrate preparation and epitaxial growth. P.S. performed the low-temperature STM characterization. J.-R.J., A.K.P., J.M.T., and F.R. assisted with the XMCD measurements in BESSY; P.G. and M.V. assisted and performed part of the XAS/XMCD experiments in ALBA. All authors discussed the data and commented on the manuscript. S.S.P.P. supervised the entire project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data discussed in the main text or the supplementary materials are uploaded in Zenodo (*57*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abd5146
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1
References (*58*–*67*)

25 June 2020; resubmitted 6 February 2021
Accepted 15 September 2021
10.1126/science.abd5146

PEROVSKITES

Liquid-phase sintering of lead halide perovskites and metal-organic framework glasses

Jingwei Hou[1]*, Peng Chen[1,2], Atul Shukla[3,4], Andraž Krajnc[5], Tiesheng Wang[6], Xuemei Li[1], Rana Doasa[7], Luiz H. G. Tizei[8], Bun Chan[9], Duncan N. Johnstone[10], Rijia Lin[1], Tobias U. Schülli[11], Isaac Martens[11], Dominique Appadoo[12], Mark S' Ari[7], Zhiliang Wang[1,2], Tong Wei[13], Shih-Chun Lo[4,14], Mingyuan Lu[15], Shichun Li[16], Ebinazar B. Namdas[3,4], Gregor Mali[5], Anthony K. Cheetham[17,18], Sean M. Collins[19]*, Vicki Chen[1], Lianzhou Wang[1,2]*, Thomas D. Bennett[10]*

Lead halide perovskite (LHP) semiconductors show exceptional optoelectronic properties. Barriers for their applications, however, lie in their polymorphism, instability to polar solvents, phase segregation, and susceptibility to the leaching of lead ions. We report a family of scalable composites fabricated through liquid-phase sintering of LHPs and metal-organic framework glasses. The glass acts as a matrix for LHPs, effectively stabilizing nonequilibrium perovskite phases through interfacial interactions. These interactions also passivate LHP surface defects and impart bright, narrow-band photoluminescence with a wide gamut for creating white light-emitting diodes (LEDs). The processable composites show high stability against immersion in water and organic solvents as well as exposure to heat, light, air, and ambient humidity. These properties, together with their lead self-sequestration capability, can enable breakthrough applications for LHPs.

Lead halide perovskites (LHPs) exhibit tunable bandgaps, high charge carrier mobilities, and bright narrow-band photoluminescence (PL) that could offer advantages for optoelectronic applications over conventional silicon (Si)–based and binary II-VI, III-V, and IV-VI semiconducting materials (*1*). However, for successful technological integration, LHPs must overcome their inherent polymorphism; decomposition upon exposure to polar solvents, oxygen, heat, and light; the presence of trap states; and the phase segregation and leaching of toxic heavy metal ions (*2*, *3*). Targeted high optical absorptivity and direct band gaps optimal for photovoltaics and red-light light-emitting diode (LEDs), for example, are found in the $CsPbI_3$ pseudo-cubic "black" phases (α-, β-, and γ-phases), but thermodynamic factors promote their conversion to the inactive non-perovskite "yellow" δ-phase under ambient conditions (Fig. 1A) (*4*). LHP materials for white-light LEDs will critically depend on stabilization of this red emitter, ideally combined in a single broad-band luminescent material architecture.

The formation of LHP composites may offer solutions to some of these problems (*5*), but the ionic nature of LHPs is not entirely conducive to composite fabrication. Functional penalties incurred include LHP aggregation and decomposition, poor mechanical stability caused by weak interfacial interactions with the chosen matrix, and the formation of high concentrations of trap states (*6*). Research into a subfamily of metal-organic frameworks (MOFs) called zeolitic imidazolate frameworks (ZIFs) has enabled access to high-temperature ZIF liquids and microporous glasses after quenching (*7*). ZIF glasses have distinct physicochemical properties in terms of their porosity, reactivity, mechanical rigidity and ductility, and optical response (*8*–*10*) and have been used as host matrices for crystalline MOFs (*11*, *12*). Together, these properties make ZIF glasses prime candidates for addressing the multiple challenges for LHP composite formation.

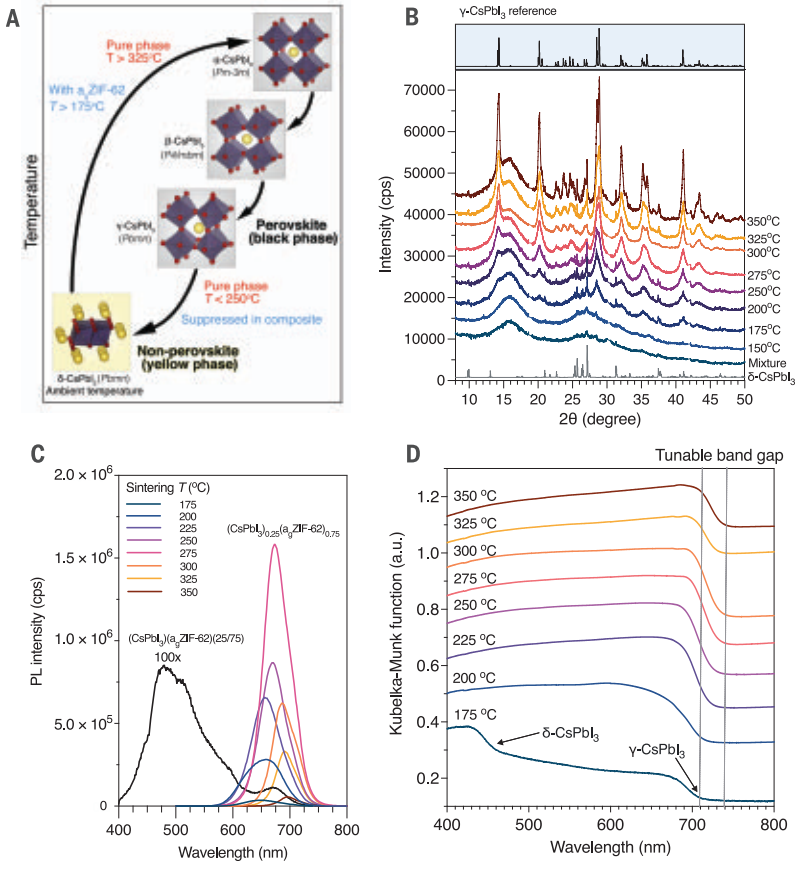


Fig. 1. Fabrication of $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ composites at various sintering temperatures. (**A**) Phase transition of $CsPbI_3$ in its pure phase and within the composites. (**B**) Ex situ room-temperature synchrotron powder XRD for $(CsPbI_3)(a_gZIF\text{-}62)(25/75)$ (marked as Mixture) and $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ composites fabricated with different sintering temperatures. cps, counts per second. (**C**) PL spectra and (**D**) ultraviolet-visible (UV-Vis) absorption spectra for $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ composites fabricated at different sintering temperatures. Arrows indicate two band edges attributed to δ- and γ-$CsPbI_3$ observed for the sample prepared at 175°C. a.u., arbitrary units.

We describe a new class of composites, fabricated by means of liquid-phase sintering of crystalline LHPs and ZIF glass matrices, and show that industrial powder processing techniques used to form high-performance composites can be applied to chemically dissimilar LHPs and ZIF glasses. ZIF-62 {$Zn[(Im)_{1.95}(bIm)_{0.05}]$} (Im, imidazolate; bIm, benzimidazolate) and $CsPbI_3$ were first synthesized mechanochemically and showed the expected phase transitions (Fig. 1A and figs. S1 to S3) (*13*). We then mixed 25 wt % $CsPbI_3$ with ZIF-62 glass [denoted as a_gZIF-62, glass transition temperature (T_g) ~ 304°C] (fig. S3B), and the mixtures are termed $(CsPbI_3)(a_g$ZIF-62$)(25/75)$ (*11*). The ex situ synchrotron powder x-ray diffraction (XRD) pattern of $(CsPbI_3)(a_g$ZIF-62$)(25/75)$ (Fig. 1B, mixture pattern) exhibited weak Bragg peaks, ascribed to the nonperovskite δ-$CsPbI_3$ phase. The mixture was sintered at different temperatures (up to 350°C) and then quenched with liquid nitrogen (referred to as cryogenic quenching) under flowing argon (Ar). The resultant composites, called $(CsPbI_3)_{0.25}(a_g$ZIF-62$)_{0.75}$, showed XRD features consistent with the metastable γ-$CsPbI_3$ phase, with gradually increasing intensity with higher sintering temperatures (Fig. 1B and fig. S4). Negligible weight loss was observed during sintering (fig. S5).

The broad PL emission of a_gZIF-62 was reduced after mixing with $CsPbI_3$, which we attribute to photon absorption by $CsPbI_3$ (fig. S6) (*14*). The $(CsPbI_3)_{0.25}(a_g$ZIF-62$)_{0.75}$ composites started to show red PL emission after sintering-quenching at 175°C, with the strongest PL obtained with 275°C (Fig. 1C). Higher sintering temperatures red-shifted the PL maxima (fig. S7), which was concomitant with an observed decrease in the optical band gaps (Fig. 1D and fig. S6C). They also led to a lower defect density and enhanced homogeneity for the $CsPbI_3$ component, as indicated by the reduced PL full-widths at half-maximum (FWHM) and the

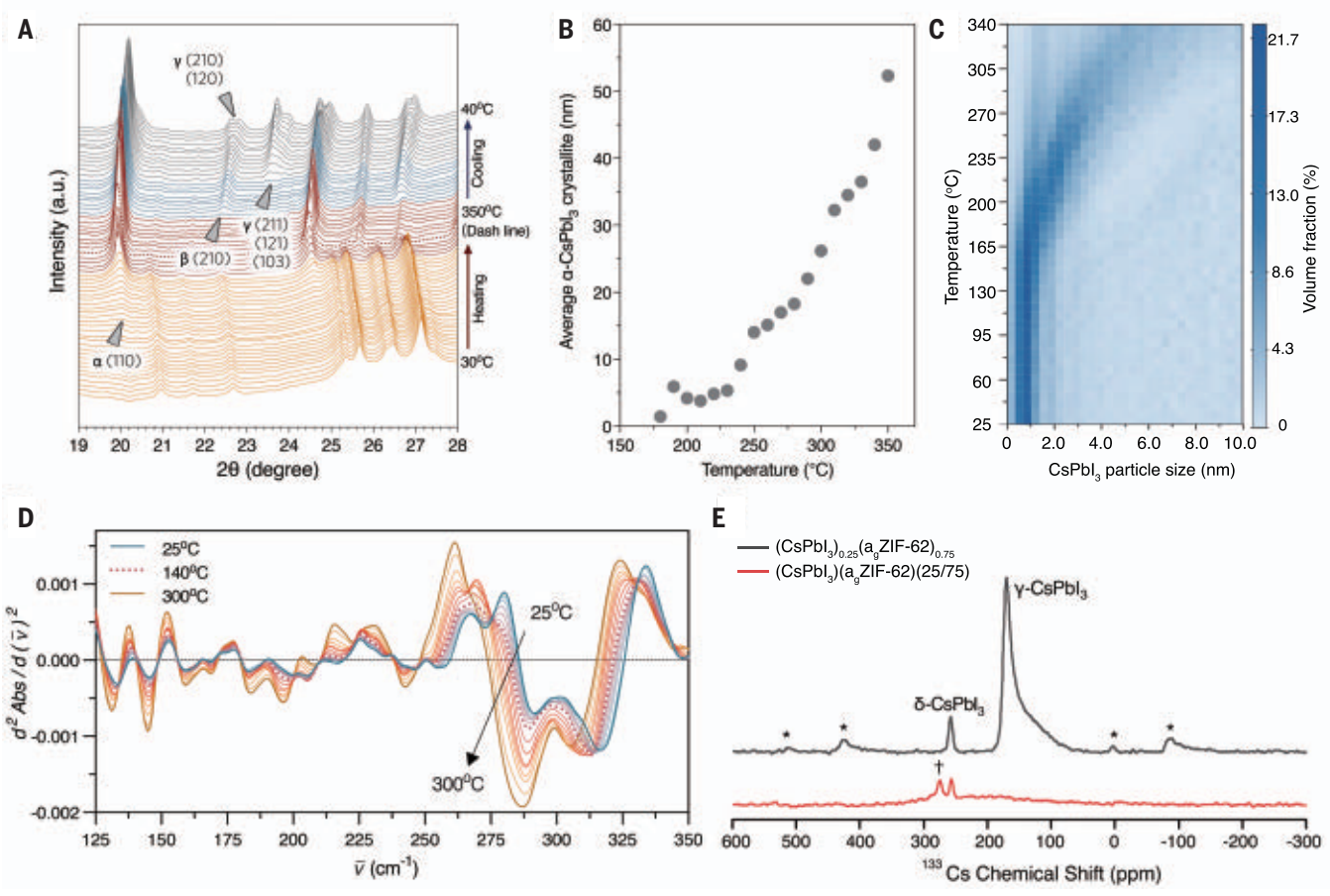


Fig. 2. Structure and bonding evolution during sintering. (**A**) Temperature-resolved, high-resolution in situ synchrotron powder XRD for $(CsPbI_3)(a_g$ZIF-62$)(25/75)$, with the Bragg peak *hkl* indices marked for different $CsPbI_3$ phases. The dominating phases are color-coded as δ, yellow; α, red; β, blue; and γ, gray. (**B**) Average sizes of α-$CsPbI_3$ deconvoluted from in situ powder XRD. (**C**) $CsPbI_3$ particle-size evolution during sintering fitted from in situ SAXS patterns. (**D**) Temperature-resolved second-derivative in situ THz FarIR spectra for $(CsPbI_3)(a_g$ZIF-62$)(25/75)$ during the first heating ramp. (**E**) ^{133}Cs MAS NMR spectra of $(CsPbI_3)(a_g$ZIF-62$)(25/75)$ and 275°C sintered $(CsPbI_3)_{0.25}(a_g$ZIF-62$)_{0.75}$. Asterisks indicate spinning sidebands, and dagger symbol (†) indicates the weak signal of CsI.

[1]School of Chemical Engineering, The University of Queensland, St Lucia, QLD, 4072 Australia. [2]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St Lucia, QLD, 4072 Australia. [3]School of Mathematics and Physics, The University of Queensland, St Lucia, QLD, 4072 Australia. [4]Centre for Organic Photonics and Electronics, The University of Queensland, Brisbane, QLD 4072, Australia. [5]Department of Inorganic Chemistry and Technology, National Institute of Chemistry, 1001 Ljubljana, Slovenia. [6]School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. [7]School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, UK. [8]Université Paris-Saclay, CNRS, Laboratoire de Physique des Solides, 91405, Orsay, France. [9]Graduate School of Engineering, Nagasaki University, Nagasaki 852-8521 Japan. [10]Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, CB3 0FS, UK. [11]The European Synchrotron Radiation Facility (ESRF), 38000 Grenoble, France. [12]Australian Synchrotron, Clayton, VIC, 3168 Australia. [13]College of Science, Civil Aviation University of China, Tianjin 300300, China. [14]School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD, 4072 Australia. [15]School of Mechanical and Mining Engineering, The University of Queensland, St Lucia, QLD, 4072 Australia. [16]Institute of Chemical Materials, China Academy of Engineering Physics, Mianyang 621900, China. [17]Materials Research Laboratory, University of California, Santa Barbara, CA 93106, USA. [18]Department of Materials Science and Engineering, National University of Singapore, Singapore, 117576 Singapore. [19]School of Chemical and Process Engineering and School of Chemistry, University of Leeds, Leeds LS2 9JT, UK.
***Corresponding author. Email: jingwei.hou@uq.edu.au (J.H.); s.m.collins@leeds.ac.uk (S.M.C.); l.wang@uq.edu.au (L.W.); tdb35@cam.ac.uk (T.D.B.)**

longer excited-state lifetimes (fig. S8 and table S1) (*15*). Compared with a slower quenching, rapid cryogenic quenching formed materials with optimal PL lifetimes and PL quantum yields (PLQYs) (>50%) (figs. S9 and S10 and table S1).

Temperature-resolved high-resolution in situ synchrotron powder XRD was collected for $(CsPbI_3)(a_gZIF\text{-}62)(25/75)$ (Fig. 2A and fig. S11). The emerging peaks from ~170°C indicate the formation of α-$CsPbI_3$ (*Pm-3m*). These peaks intensified at higher sintering temperatures. During the quenching stage, the gradual emergence of β-$CsPbI_3$ (*P4/mbm*) (from ~250°C) and γ-$CsPbI_3$ (*Pbnm*) (from ~150°C) was evidenced (*16*). The deconvoluted α-$CsPbI_3$ crystallite size increased during sintering (Fig. 2B), which is consistent with the changes in band gap caused by quantum-confinement effects (*5*).

The evolution of α-$CsPbI_3$ crystallite size can be attributed to coarsening of $CsPbI_3$ grains and the phase transition from bulkier δ-$CsPbI_3$ crystallites, a cascade confirmed with synchrotron in situ small-angle x-ray scattering (SAXS). Coarsening of $CsPbI_3$ grains mainly occurred in the size range smaller than the XRD deconvoluted crystallite size at <10 nm, starting from 165°C (Fig. 2C and figs. S12 to S14). Upon sintering, atoms in $CsPbI_3$ grains became mobile from the Tamman temperature (T_{Tamman} ~ 103°C) as approximated by 0.5 T_{melt} (melt temperature) in degrees kelvin (*16*, *17*). A similar response could also be expected for a_gZIF-62. Characteristic of liquid-phase sintering, $CsPbI_3$ grain coarsening and composite densification were observed at a temperature lower than the inherent T_g of a_gZIF-62 (~304°C) (fig. S15) (*18*).

The emergence of an interface resulting from densification occurs analogously to surface energy–controlled transitions from δ- to α-phase in solvent-modulated or ligand-capped $CsPbI_3$ quantum dots (*19*, *20*), with the interfacial energy dominant for smaller grains resulting in phase transitions at lower temperatures. To examine our hypothesis that intimate interfacial contact is critical for phase control, we synthesized $[Zn(Im)_{1.75}(bIm)_{0.25}]$ a_gZIF-62 with a higher T_{melt} and higher viscosity caused by bulkier bIm ligands and subsequently demonstrated the expected higher residual δ-$CsPbI_3$ content in the composite (fig. S16).

We further probed the changes in interfacial bonding within $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ by means of temperature-resolved synchrotron terahertz (THz) radiation and far-infrared (FarIR)

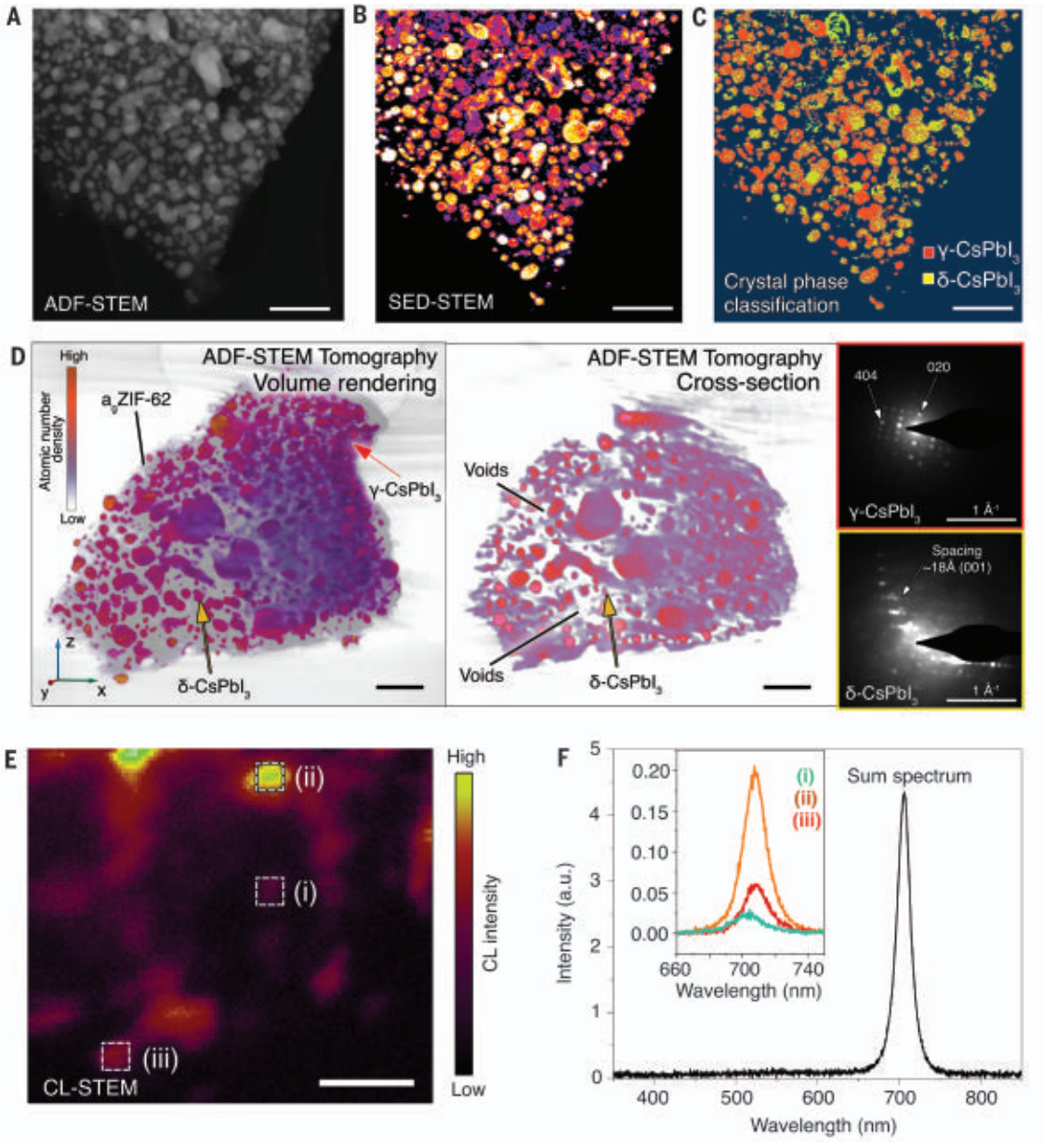


Fig. 3. Phase distribution for the $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ composite fabricated with 300°C sintering. (**A**) ADF-STEM image, (**B**) SED-STEM mapping, and (**C**) $CsPbI_3$ crystal phase classification results for $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ composite. (**D**) Volume rendering of a tomographic reconstruction of $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ and a single cross-sectional plane extracted from the volume. Color-coded arrows indicate the regions where electron diffraction data were collected. Scale bars, (A) to (D) 250 nm. (**E**) CL-STEM mapping of the integrated CL intensity. Scale bar, 70 nm. (**F**) CL spectra acquired at each STEM probe position. (Inset) The sum CL spectrum of the whole region in (E).

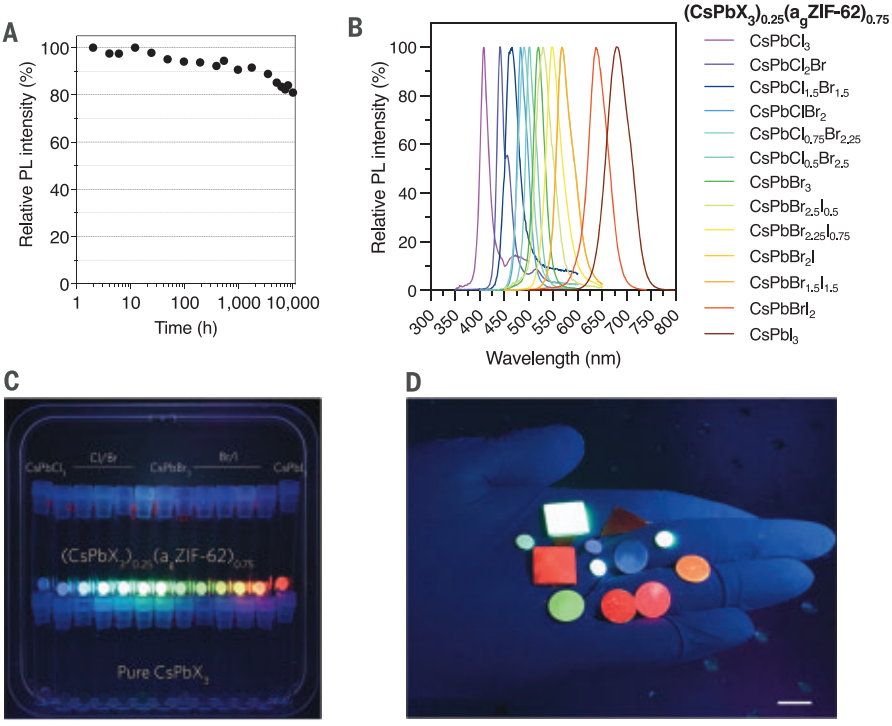


Fig. 4. Stability and optical performance of the composites. (**A**) Change of the relative PL intensity for $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ immersed in the Milli-Q water. Sample was sintered at 300°C. (**B**) Normalized PL intensities of the $(CsPbX_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ composites (X = Cl, Br, I, and mixed halide ions). (**C** and **D**) Optical photos of the composites and pure $CsPbX_3$ under 365 nm UV light. Composites for (B) to (D) were sintered at 275°C. Scale bar, 1 cm.

vibrational spectroscopy. The second-derivative spectra revealed the fine vibrational modes of Zn tetrahedra (Fig. 2D and figs. S17 and S18) (*9*). The modes assigned to Zn-N vibrations (~287 cm^{-1}) and Zn-I stretching (~135 cm^{-1}) within $Zn(Im)_2(bIm)I$ tetrahedra through density functional theory (fig. S19) began to intensify with increasing temperature from ~140°C. These changes were consistent with the endothermic response at ~140°C in the first heating ramp of differential scanning calorimetry (DSC) measurements, which was concomitant with changes in $CsPbI_3$ binding observed in phonon signatures and in ex situ spectroscopy (figs. S20 to S23).

Magic-angle–spinning nuclear magnetic resonance (MAS NMR) spectroscopy provided insights into the different species complementary to vibrational spectroscopy. The broader signals from composite ^{13}C and ^{15}N spectra indicated additional disorder of ZIF component over the powder mixture (fig. S24). The ^{133}Cs MAS NMR spectra of $(CsPbI_3)(a_gZIF\text{-}62)(25/75)$ exhibited narrow signals of δ-$CsPbI_3$ [260 parts per million (ppm)] (*21*) and CsI (~280 ppm). Broad, low signals extending between 0 and ~350 ppm can be ascribed to poorly crystalline, highly defective $CsPbI_3$ (Fig. 2E and fig. S25). After sintering, the broad contributions and CsI peaks diminished, and the major signals stemmed from γ-$CsPbI_3$. These signals exhibited notable shoulders (160 to 80 ppm), with shoulder intensities highly dependent on the sintering conditions. They could be assigned to Cs nuclei on or near the surface of $CsPbI_3$ grains where structural defects, sites of the interaction between the γ-$CsPbI_3$ and $a_gZIF\text{-}62$, or both are abundant. Also, the signals of the δ-$CsPbI_3$ of the same sample exhibited no shoulder, which is consistent with less interfacial contact between δ-$CsPbI_3$ and $a_gZIF\text{-}62$.

These observations allowed us to propose a mechanism for γ-$CsPbI_3$ stabilization within composites (*22*, *23*). The α-, β-, and γ-phases of $CsPbI_3$ have double-well phonon modes at the center of the Brillouin zone, driving the phase transition to δ-$CsPbI_3$ in a concerted phonon manner (*24*). The interfacial bonding disrupts the local Pb-I sublattice phonon modes and therefore avoids the harmonic order-disorder entropy (*25*, *26*). Together with the physical confinement effect offered by the matrices, these factors counter the strong thermodynamic driving force to form δ-$CsPbI_3$.

We evaluated this mechanism further and verified embedded nanocrystals of γ-$CsPbI_3$ as the source of luminescence using microscopic measurements. After sintering, the mixture of particles became a monolith, with a smooth surface observed in scanning electron microscopy (SEM) (figs. S26 and S27). Annular dark-field scanning transmission electron microscopy (ADF-STEM) of $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ showed pronounced atomic number contrast between the two phases, which was further corroborated with energy-dispersive x-ray spectroscopy (STEM-EDS) elemental distribution mapping (Fig. 3A and fig. S28). The crystalline and amorphous regions were identified by means of scanning electron diffraction (SED) (*27*), with regions exhibiting Bragg diffraction corresponding to crystalline $CsPbI_3$ grains (Fig. 3B). Convolutional neural network (CNN) classification identified γ-$CsPbI_3$ as the major phase within the composite fragment. Individual grains were single-crystalline, whereas the speckle in the classification map arose from inherent ambiguities because of overlap in the diffraction peaks expected from δ- and γ-$CsPbI_3$ (Fig. 3C and figs. S29 and S30). The average size of $CsPbI_3$ from STEM was ~30 nm (fig. S31), readily modulated by extended ball milling before sintering, which further enhanced the composite PLQY to >65% because of a more pronounced quantum confinement effect (fig. S32) (*5*).

To probe the internal structure, we performed ADF-STEM tomography on a shard (>1 μm) of $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ (Fig. 3D). The voids in cross sections of the volume are characteristic of densification processes in liquid phase sintering (*18*). Point diffraction data identified both δ-$CsPbI_3$ and γ-$CsPbI_3$ within the particle. A high degree of interfacial contact was correlated with γ-$CsPbI_3$, which is consistent with the hypothesized phase control through interfacial stabilization (fig. S33). STEM-based cathodoluminescence (CL) detected strong, narrow luminescence from isolated grains (<40 nm), with minor interparticle emission wavelength shifts (Fig. 3, E and F, and fig. S34). The variation of CL intensity is complex in origin, highly sensitive to the crystal quality and exposure to unpassivated surface states and particle size effects (*28*). Despite this, the CL spectra from individual grains provided incontrovertible evidence of luminescence from glass-bound nanocrystals of γ-$CsPbI_3$.

Returning to the aim of achieving long device lifetimes, we evaluated the composites in diverse environmental and operational settings. The rigid, hydrophobic $a_gZIF\text{-}62$ provided protection for $CsPbI_3$ (figs. S35 and S36), leading to stable PL emission for $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ after extended (~20 hours) sonication in various nonpolar, polar protic, and polar aprotic organic solvents (fig. S37). The composite also exhibited stability against 10,000 hours immersion in water, storage under ambient conditions for 650 days, mild heating, and continuous laser excitation (~57 mW/cm^2) for >5000 s (Fig. 4A and figs. S38 to S40). The microporous composite design presents a key route to sequestration of toxic components (figs. S41 and S42), or to potential photochemical platforms where the $CsPbI_3$ crystals are not electronically insulated (fig. S43). $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ made from mechanochemical precursors have similar performance compared with

the solvothermal precursors, making the composite promising for up-scaling (fig. S44). Collectively, $(CsPbI_3)_{0.25}(a_gZIF\text{-}62)_{0.75}$ offers substantial advantages over LHP composites with other substrates (fig. S45 and table S2).

Last, an array of composites were formed from $CsPbX_3$ (X = Cl, Br, and mixed halide ions) and a_gZIF-62, showing a wide color gamut with narrow PL peaks (Fig. 4, B and C, and table S3). For all the $CsPbX_3$ composites, their absolute PL intensities were at least two orders of magnitude higher than those of the corresponding pure $CsPbX_3$ samples, either as synthesized or after being treated with identical sintering (fig. S46). These properties, together with the high processibility (Fig. 4D), render these monolithic materials ideal candidates for downshifting white LEDs (fig. S47).

REFERENCES AND NOTES

1. M. V. Kovalenko, L. Protesescu, M. I. Bodnarchuk, *Science* **358**, 745–750 (2017).
2. X. Li *et al*., *Nature* **578**, 555–558 (2020).
3. T. A. S. Doherty *et al*., *Nature* **580**, 360–366 (2020).
4. Y. Wang *et al*., *Science* **365**, 591–595 (2019).
5. J. Shamsi, A. S. Urban, M. Imran, L. De Trizio, L. Manna, *Chem. Rev.* **119**, 3296–3348 (2019).
6. Q. A. Akkerman, G. Rainò, M. V. Kovalenko, L. Manna, *Nat. Mater.* **17**, 394–405 (2018).
7. T. D. Bennett *et al*., *J. Am. Chem. Soc.* **138**, 3484–3492 (2016).
8. R. S. K. Madsen *et al*., *Science* **367**, 1473–1476 (2020).
9. J. Hou *et al*., *J. Am. Chem. Soc.* **142**, 3880–3890 (2020).
10. L. Frentzel-Beyme, M. Kloß, P. Kolodzeiski, R. Pallach, S. Henke, *J. Am. Chem. Soc.* **141**, 12362–12371 (2019).
11. J. Hou *et al*., *Nat. Commun.* **10**, 2580 (2019).
12. C. W. Ashling *et al*., *J. Am. Chem. Soc.* **141**, 15641–15648 (2019).
13. J. A. Steele *et al*., *Science* **365**, 679–684 (2019).
14. N. J. L. K. Davis *et al*., *J. Phys. Chem. C Nanometer Interfaces* **121**, 3790–3796 (2017).
15. K. M. Boopathi *et al*., *ACS Energy Lett.* **5**, 642–649 (2020).
16. A. Marronnier *et al*., *ACS Nano* **12**, 3477–3486 (2018).
17. Y. Dai, P. Lu, Z. Cao, C. T. Campbell, Y. Xia, *Chem. Soc. Rev.* **47**, 4314–4331 (2018).
18. R. M. German, P. Suri, S. J. Park, *J. Mater. Sci.* **44**, 1–39 (2009).
19. Y. Wang *et al*., *Adv. Mater.* **32**, e2000186 (2020).
20. B. Zhao *et al*., *J. Am. Chem. Soc.* **140**, 11716–11725 (2018).
21. A. Karmakar *et al*., *Chem. Commun.* **55**, 5079–5082 (2019).
22. R. Wang *et al*., *Science* **366**, 1509–1513 (2019).
23. S. Bai *et al*., *Nature* **571**, 245–250 (2019).
24. A. Marronnier *et al*., *J. Phys. Chem. Lett.* **8**, 2659–2665 (2017).
25. M. Grechko, S. A. Bretschneider, L. Vietze, H. Kim, M. Bonn, *Angew. Chem. Int. Ed.* **57**, 13657–13661 (2018).
26. B. Wang, N. Novendra, A. Navrotsky, *J. Am. Chem. Soc.* **141**, 14501–14504 (2019).
27. O. Panova *et al*., *Micron* **88**, 30–36 (2016).
28. J. T. Griffiths *et al*., *ACS Appl. Energy Mater.* **2**, 6998–7004 (2019).

ACKNOWLEDGMENTS

We acknowledge D. D'Alessandro and Y. Sun for discussions. **Funding:** This work was supported by the Australian Research Council (DE190100803, DE210100930, DP180103874, DE190101152, DP200101900, and FL190100139); Department of Industry, Innovation and Science (AISRF53765); University of Queensland (UQECR2057677); Australian Centre for Advanced Photovoltaics fellowship and Australian Renewable Energy Agency; Henry Royce Institute for a summer undergraduate internship; National Natural Science Foundation of China (51772326); RIKEN Information Systems and Cybersecurity (Project Q20266); Slovenian Research Agency (research core funding P1-0021); Ras al Khaimah Center for Advanced Materials; Royal Society and Leverhulme Trust for a University Research Fellowship (UF150021) and Philip Leverhulme Prize (2019); European Union's Horizon 2020 research and innovation program (823717–ESTEEM3); and the National Agency for Research future investment TEMPOS-CHROMATEM (ANR-10-EQPX-50). Part of this research was undertaken on the THz/Far-IR, SAXS, and PD beamlines at the Australian Synchrotron, part of ANSTO (M15988 and M15433). The authors acknowledge the Centre for Microscopy and Microanalysis and the Australian National Fabrication Facility, the University of Queensland; the Leeds EPSRC Nanoscience and Nanotechnology Facility (LENNF); ESRF / ID31 beamline; and the Diamond Light Source for access and support in the use of the electron Physical Sciences Imaging Centre (MG21980 and MG25140). **Author contributions:** Conceptualization: J.H. and T.D.B. Methodology: J.H., P.C., T.W., S.M.C., Z.W., and T.U.S. Investigation: J.H., A.S., S.-C.L., T.W., E.B.N., A.K., G.M., R.D., L.H.G.T., D.N.J., M.S'A., S.M.C., B.C., X.L., R.L., S.L., M.L., I.M., and D.A. Funding acquisition: J.H., S.M.C., G.M., A.K.C., V.C., L.W., and T.D.B. Project administration: J.H., S.M.C., L.W. Writing, original draft: J.H., A.K.C., V.C., L.W., and T.D.B. Writing, review and editing: all authors. **Competing interests:**. J.W., V.C., and L.W. are inventors on Australian Provisional Patent Application no. 2021902824 held by the University of Queensland, titled "Composite Material" and related to composite glass material and methods of making composite glass materials. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. The composite glass samples are available from J.W. and L.W. under a materials transfer agreement with the University of Queensland.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf4460
Materials and Methods
Supplementary Text
Figs. S1 to S47
Tables S1 to S3
References (*29*–*61*)

27 October 2020; resubmitted 28 May 2021
Accepted 17 September 2021
10.1126/science.abf4460

PLANT SCIENCE

NIN-like protein transcription factors regulate leghemoglobin genes in legume nodules

Suyu Jiang[1], Marie-Françoise Jardinaud[2], Jinpeng Gao[1], Yann Pecrix[2,3], Jiangqi Wen[4], Kirankumar Mysore[4], Ping Xu[5], Carmen Sanchez-Canizares[6], Yiting Ruan[1], Qiujiu Li[1], Meijun Zhu[1], Fuyu Li[1], Ertao Wang[1], Phillip S. Poole[6], Pascal Gamas[2], Jeremy D. Murray[1,7]*

Leghemoglobins enable the endosymbiotic fixation of molecular nitrogen (N_2) in legume nodules by channeling O_2 for bacterial respiration while maintaining a micro-oxic environment to protect O_2-sensitive nitrogenase. We found that the NIN-like protein (NLP) transcription factors NLP2 and NIN directly activate the expression of leghemoglobins through a promoter motif, resembling a "double" version of the nitrate-responsive elements (NREs) targeted by other NLPs, that has conserved orientation and position across legumes. CRISPR knockout of the NRE-like element resulted in strongly decreased expression of the associated leghemoglobin. Our findings indicate that the origins of the NLP-leghemoglobin module for O_2 buffering in nodules can be traced to an ancient pairing of NLPs with nonsymbiotic hemoglobins that function in hypoxia.

Most legume species can form an endosymbiosis with rhizobia bacteria, forming highly specialized root nodules containing cells in which rhizobial nitrogenase reduces dinitrogen to ammonia. Nitrogenase is rapidly and irreversibly deactivated by oxygen, whereas N_2 fixation is energy-intensive, so the rate of N_2 fixation is a compromise between the O_2 required to support the energy demands of the symbionts and the poisoning of nitrogenase by O_2, the so-called "oxygen paradox" (*1*). To address this constraint, up to 40% of nodule protein consists of leghemoglobins (LgHbs), which contain an O_2-binding heme group that colors the nodule red. LgHbs feature an extremely fast O_2 association rate and a relatively slow O_2 dissociation rate, allowing them to buffer the free oxygen concentration at ~10 nM and thereby maintain O_2 levels compatible with N_2 fixation. Notwithstanding the central role of LgHbs in N_2 fixation, their transcriptional regulation and the manner of their recruitment to nodulation have not been elucidated.

The founding member of the NIN-like protein family, NIN (Nodule Inception), is required for infection by rhizobia as well as for nodule formation (*2*). In contrast, *Medicago truncatula* NLP1 and *Lotus japonicus* NLP4 are required for nitrate suppression of nodulation (*3*, *4*). *NIN* and *NLP2*, however, are the only NLPs expressed in medicago nodules (Fig. 1A and fig. S1A). The expression of *NLP2* and several *LgHb*s was strongly reduced in nodules induced by the rhizobia mutants *bacA* and *fixJ*; in these mutants, N_2 fixation is defective as a result of lack of symbiosomes or nitrogenase activity, respectively, indicating a tight connection of *NLP2* with N_2 fixation (Fig. 1B).

To investigate NLP2's role, we used an aeroponic system to evaluate nodulation on *NLP2*-knockout roots. The resultant nodules had normal zonation but were smaller and had reduced N_2 fixation potential (Fig. 1, C and D). We then isolated two mutants, *nlp2-1* and *nlp2-2* (fig. S1B). The *nlp2-1* mutant had smaller nodules in aeroponics, matching the CRISPR knockout phenotype (Fig. 1E), and had a greatly reduced plant biomass but developed a normal number of nodules (Fig. 1F). However, when grown in substrate, the mutant developed normal-sized nodules (fig. S1C). The increase in phenotype severity seen in aeroponics versus substrate conditions may reflect the much larger and more extensively nodulated root system that develops under aeroponics, which may exaggerate the

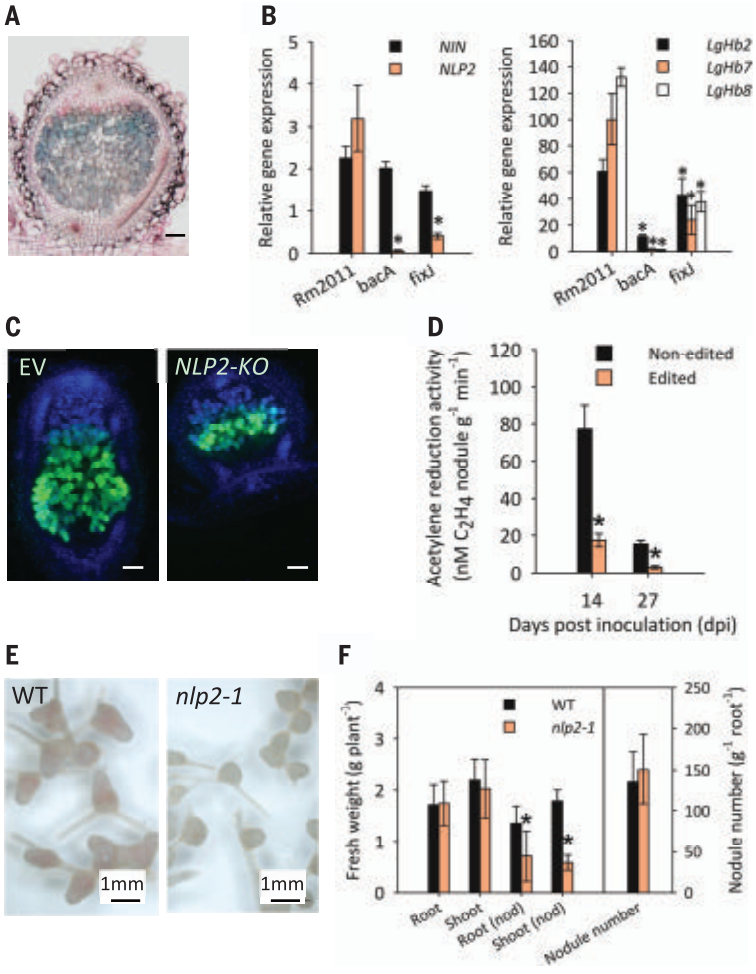


Fig. 1. Expression of *NLP2* in *M. truncatula* nodules and *nlp2* nodulation phenotypes. (**A**) Nodule expression of *NLP2*. Shown is a histochemically stained nodule cross section [15 days post-inoculation (dpi) with *S. meliloti* Rm2011] expressing *pMtNLP2:GUS*. Scale bar, 100 μm. (**B**) Expression [as measured by quantitative reverse transcription polymerase chain reaction (qRT-PCR)] of *NLP2*, *NIN*, *LgHb2*, *LgHb7*, and *LgHb8* relative to reference genes in nodules of wild-type (WT) plants inoculated with *S. meliloti* Rm2011 grown in attapulgite substrate with no nitrogen. $n = 3$. (**C**) Representative DAPI-stained 70-μm nodule sections of *NLP2* knockout (*NLP2*-KO) and control (EV) roots grown using aeroponics with no nitrogen. Scale bar, 100 μm. Nodules were evaluated 11 dpi with *S. meliloti* Rm2011 carrying a *pNifH:GFP* reporter. (**D**) Acetylene reduction activity of nodules of *NLP2* CRISPR-edited roots grown aeroponically without nitrogen. WT *M. truncatula* seedlings (A17) were transformed with a gRNA targeting *NLP2* and roots were genotyped. Mean numbers of nodules from edited and non-edited roots were compared ($n = 19$) with a Kruskal nonparametric test, $\alpha = 0.05$. (**E**) Nodules of *nlp2-1* and WT nodules grown aeroponically without nitrogen, 18 dpi with *S. meliloti* Rm2011. (**F**) Left: Biomass of nodulated (nod) (18 dpi with *S. meliloti* Rm2011) and uninoculated plants. Right: Nodule number. Plants were grown aeroponically without nitrogen; $n = 10$ for WT, 11 for *nlp2-1*. In (B), (D), and (F), error bars denote SD. In (B) and (F), means were compared with Student's *t* test, $\alpha = 0.05$. In (B), (D), and (F), $^*P < 0.05$.

N-fixation defect in *nlp2*. Nonetheless, the N-fixation potential of both *nlp2-1* and *nlp2-2* was also compromised, albeit to a lesser degree, when grown in substrate (fig. S1D) and could be partially complemented by expression of *NLP2* from its native promoter, but not by a mutant version of *NLP2* (fig. S1, E and F). Furthermore, $^{15}N_2$ assimilation was compromised in *nlp2-1* plants nodulated in substrate (Fig. 2A), which corresponded to a decrease in N content in nodules and shoot, while %N was unchanged (Fig. 2A and fig. S2D). Evaluation of *nlp2* mutants under inhibitory nitrate regimes indicated that loss of NLP2 does not affect N suppression of nodulation (fig. S2, A to C). Together these results suggest a positive role for NLP2 in N_2 fixation.

[1]CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Centre for Excellence in Molecular Plant Sciences (CEMPS), Shanghai Institute of Plant Physiology and Ecology (SIPPE), Chinese Academy of Sciences, Shanghai, China. [2]LIPME, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France. [3]CIRAD, UMR PVBMT, Pôle de Protection des Plantes, Saint-Pierre 97410, France. [4]Noble Research Institute, Ardmore, OK 73401, USA. [5]Shanghai Engineering Research Center of Plant Germplasm Resource, College of Life Sciences, Shanghai Normal University, Shanghai 200234, China. [6]Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK. [7]John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK.
***Corresponding author. Email: jeremy.murray@jic.ac.uk**

In addition to reduced N fixation, *nlp2* nodules appeared less pink (fig. S2E); this was reflected in lower LgHb expression and heme levels (Fig. 2, B and C, and fig. S2F). We then profiled the gene expression of *nlp2-1* nodules under different N regimes using substrate where the mutant forms normal-sized nodules, rather than using aeroponics where the smaller nodules on the mutant would confound the analysis. The results revealed nitrate-dependent effects on gene expression (Fig. 2D and data S1). Genes encoding LgHbs, heme biosynthesis enzymes, and nitrate uptake transporters were strongly down-regulated in *nlp2-1* (Fig. 2, E and F, and table S1), whereas jasmonate-related genes were up-regulated (table S2), as previously reported in root hairs of *M. truncatula nin* (*5*). Expression of *NLP2* from its own promoter partially rescued gene expression in *nlp2*, but a mutant version did not (Fig. 2G).

In silico promoter analysis revealed that 28 of the 55 genes with the largest overall decrease in expression in *nlp2* nodules contained nitrate-responsive elements (NREs), which are required for transcriptional activation by NLPs (*6*). This included *SymCRK* and *SEN1*, which are required for N fixation, and many genes encoding nodule cysteine-rich peptides (NCRs) (*7*, *8*) (table S3). In addition, a larger motif consisting of two partly overlapping NREs was detected in the *LgHb* genes. We designated these as "double-NREs" (dNREs), in which the second half of the first semi-palindrome (NRE1) overlaps with the first half of the second element (NRE2) (Fig. 3A). The dNREs all have the same orientation and are positioned between –100 and –300 bp upstream of the TSS (Fig. 3B). dNREs were also detected at the same position and orientation in promoters of *LgHb*s of soybean and common bean, indicating their conservation across papilionoid legumes (fig. S3A). Examination of dNREs in promoters of soybean and medicago *LgHb*s suggests that those strongly differing from the consensus motif have lower expression in nodules (fig. S3, B and C), implying the direct involvement of dNREs in the regulation of *LgHb* expression.

Our transcriptomic data verified that *NLP2* and *NIN* are the primary NLPs expressed in medicago nodules (figs. S1A and S4); because NIN binding sites strongly resemble NREs (*9*, *10*) (fig. S5), we speculated that both NLP2 and NIN may directly promote *LgHb* expression. Using an electrophoretic mobility shift assay (EMSA), we found that truncated NLP2

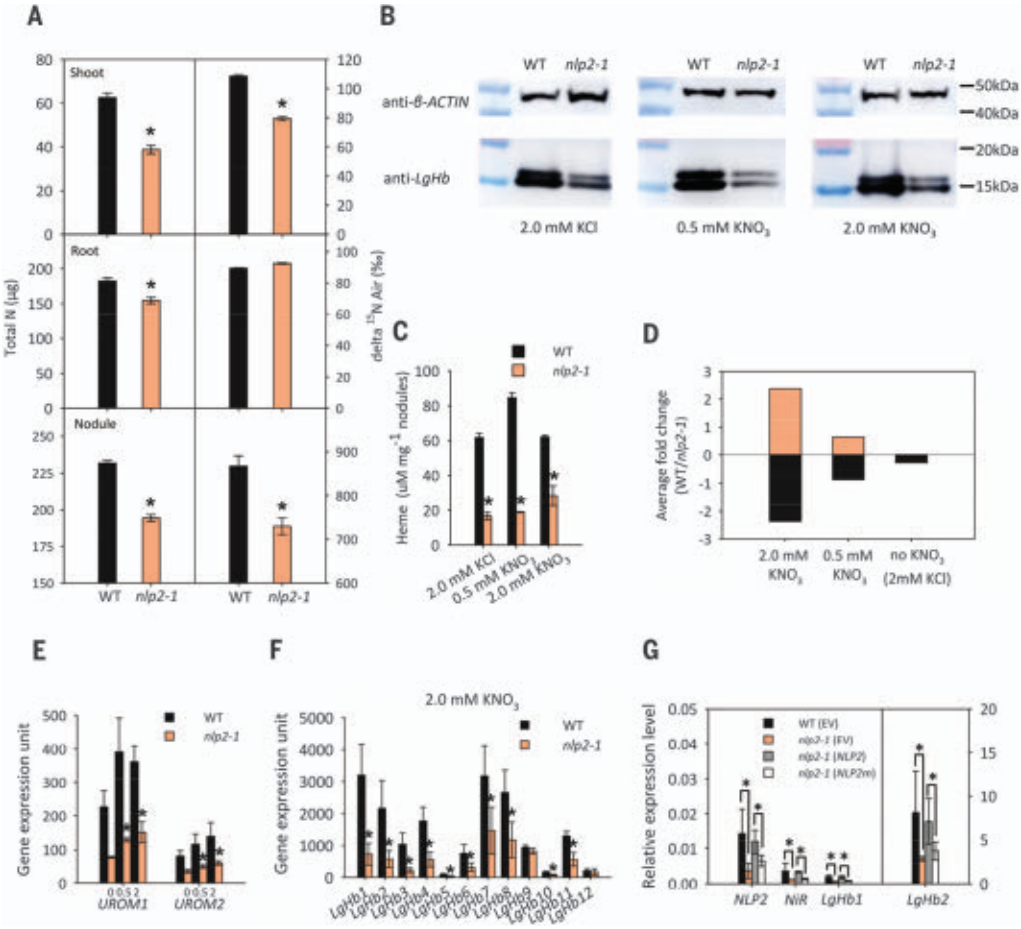


Fig. 2. Loss of NLP2 reduces nitrogen fixation, plant nitrogen content, LgHb, and heme levels. (**A**) Total N (Kjeldahl's) and N_2 fixation [$\delta^{15}N$ (‰)] in *nlp2-1* and WT nodules. Plants were nodulated under 0.5 mM KNO_3. (**B**) LgHb content of WT and *nlp2-1* nodules. Western blot uses α-LgHb antibodies to detect LgHb levels in nodules of WT and *nlp2-1* under different nitrogen regimes. The two bands reflect the presence of different-sized LgHb isoforms. (**C**) Heme content of *nlp2-1* and WT nodules under different nitrogen regimes. (**D**) Average relative (fold) changes for genes with changed expression in *nlp2-1* nodules at 2.0 mM KNO_3 (656 decreased, 915 increased) under different nitrogen treatments. (**E**) Relative expression of *UROM1/2* in *nlp2-1* and WT nodules under (left to right) 2.0 mM KCl, 0.5 mM KNO_3, and 2.0 mM KNO_3. (**F**) Expression of *LgHb*s in nodules in WT and *nlp2-1* under 2.0 mM KNO_3. (**G**) Complementation of gene expression in *nlp2-1* nodules under 2.0 mM KNO_3. Relative expression of *NLP2*, *LgHb1*, *LgHb2*, and *NiR* in *nlp2-1* nodules transformed with *pNLP2-NLP2*, *pNLP2-NLP2m*, or EV (empty vector) introduced by *A. rhizogenes* hairy-root transformation. *EF* was used as a referencegene. For (A), $n = 10$; for (C), (E), (F), and (G), $n = 3$. All plants were evaluated at 21 dpi with *S. meliloti* Rm2011. Means were compared by Student's *t* test, $\alpha = 0.05$. In (E) and (F), error bars indicate 95% confidence interval. In (A), (C), and (G), error bars denote SD. In (A), (C), and (E) to (G), $^*P < 0.05$.

and NIN proteins containing their RWP-RK DNA binding domains could bind to dNRE-containing DNA fragments but not to those with mutated dNREs (Fig. 3, C and D; details in figs. S6 and S7). To test its importance for *LgHb* expression, we mutated the dNRE element in the *LgHb1* promoter using CRISPR-Cas9. This resulted in strongly decreased expression of *LgHb1*, but not *LgHb2*, whose promoter contained the guide RNA (gRNA) target sequence but lacked the protospacer adjacent motif (PAM) (Fig. 3E and fig. S8). When expressed in *Nicotiana benthamiana* leaves, both NIN and NLP2 increased reporter activity when driven by a dNRE-containing fragment of the *LgHb7* promoter, but not when driven by a dNRE-deleted version (Fig. 3F). Versions of NIN and NLP2 with mutations in their RWP-RK domains were unable to activate the reporter (fig. S9A). Co-infiltration of *NIN* and *NLP* constructs resulted in higher reporter gene activity than either construct alone (fig. S9B). These results support a role for NLP2 and NIN in the activation of *LgHb* expression in nodules. Because *nin* loss-of-function mutants do not form nodules, we used a CRISPR-Cas9–based strategy to selectively knock out *NIN* in a tissue-specific manner by expressing Cas9 in the nodule interzone/nitrogen fixation zone of wild-type and *nlp2-1* nodules. Successfully edited nodules showed decreased *LgHb* expression, which suggests that NLP2 and NIN act redundantly to regulate the transcription of nodule *LgHb*s (fig. S10).

We then investigated whether the above regulation was a legume innovation or predated the symbiosis. Phylogenetic analysis showed that close homologs of NLP2 are found only in legumes and group separately from their next closest homologs, most of which are expressed in seeds (figs. S11 and S12), whereas NIN was reported to be present in legumes and nodulating non-legumes (*11*). The absence of NLP2 in *Prosopis alba*, a caesalpiniod legume, suggests that NLP2 arose before the divergence of papilionoid legumes. LgHbs evolved from nonsymbiotic hemoglobins (nsHbs) involved in energy maintenance under hypoxia and O_2 transport (*12*, *13*), with class I nsHbs acting in the nsHb–nitric oxide cycle (*14*, *15*). Analysis of *nsHb* promoter sequences across angiosperms revealed that they frequently harbor NRE elements, which suggests that they may be NLP targets (fig. S13, A and B). Mining of data from a genome-wide chromatin immunoprecipitation analysis revealed binding of *L. japonicus* NIN to a class I *nsHB* (*16*). In addition, *nsHB1* was identified as a direct target of AtNLP7 in an *Arabidopsis* DNA affinity purification sequencing study (*17*), and its expression was shown to be dependent on NLP7 or NLP8 in three transcriptomic studies (*18*–*20*). Our own investigation of medicago *nsHB1* showed that its expression is dependent on NLP1, and that it may be a direct target (fig. S14). These findings suggest that NLPs were recruited to nodulation as part of an ancient NLP-nsHb regulatory module. The dNRE element was detected in *Hb* promoters of a mimosoid legume, and thus appears to predate NLP2, but was absent in nodulating non-legumes (fig. S13C; these and other promoter sequences used are provided in the supplementary materials). It therefore seems likely that regulation of hemoglobins by NLPs served as a basis for recruitment of NIN in nodulating

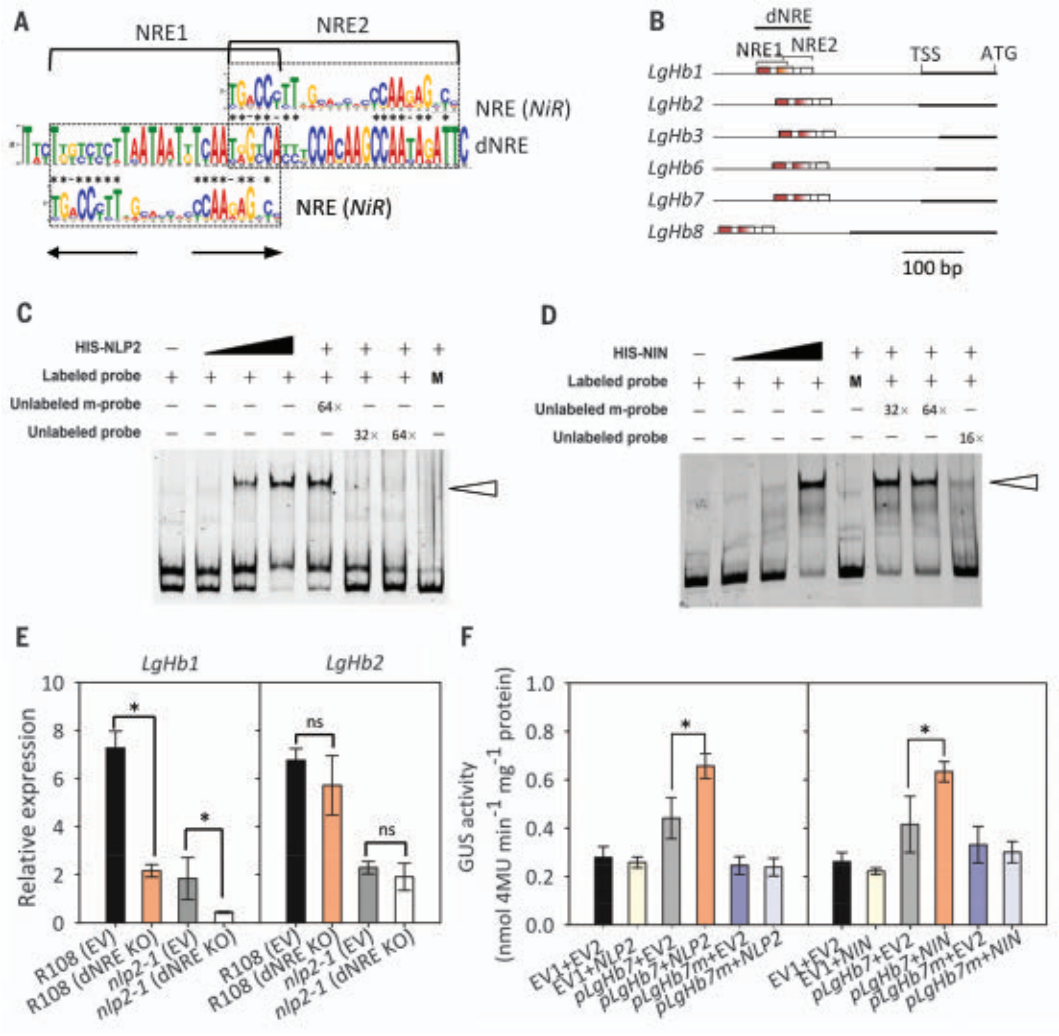


Fig. 3. NLP2 directly activates *LgHb* expression via NRE-like promoter elements. (**A**) "dNRE" motif detected in promoters of *M. truncatula nlp2-1* down-regulated genes compared with NREs detected in *NiR*s across angiosperm plants (this study). The gapped semi-palindrome is indicated by arrows. Overlapping NREs are indicated by brackets. Matches and mismatches of the NRE consensus to the dNRE are indicated by asterisks and dashes, respectively. (**B**) The position of the dNRE sequence in promoters of *M. truncatula LgHb*s (NRE1 in orange, NRE2 in white). The bold line indicates the 5′-untranslated region. TSS, transcription start site; ATG, start codon. (**C**) EMSA to test for interaction between a truncated NLP2 protein containing the RWP-RK domain with an N-terminal His-tag and the CY5-labeled *LgHb7* promoter fragment containing the dNRE. M: *LgHb7* promoter with a mutated dNRE. (**D**) EMSA to test for interaction between a truncated NIN protein containing the RWP-RK domain with an N-terminal His tag and the CY5-labeled *LgHb7* promoter fragment containing the dNRE. M: *LgHb7* promoter with deleted dNRE. (**E**) Relative expression (qRT-PCR) of *LgHb1 and LgHb2* in nodules of *M. truncatula* plants transformed with a CRISPR construct with two gRNAs targeting the dNRE of *LgHb1*. gRNA1 also matches the dNRE sequence in the *LgHb2* promoter, but there is no corresponding PAM site. The vector was introduced by hairy-root transformation; gene expression was evaluated at 21 dpi with *S. meliloti* Rm2011. *EF* was used as a reference gene. (**F**) Trans-activation assay for NLP2 and the *LgHb7* promoter driving *GUS* in *Agrobacterium tumefaciens*–co-infiltrated *Nicotiana benthamiana* leaves. EV1, empty vector no-promoter control; EV2, empty vector no-*NLP2* control; *pLgHb7*, *LgHb7* promoter fragment (−342 to −1 bp) containing the dNRE; *LgHb7m*, *LgHb7* promoter fragment with the dNRE deleted; *NIN* and *NLP2*, p35S promoter driving full-length *NLP2* or *NIN*. In (E) and (F), data are means ± SD and are compared by Student's *t* test, $\alpha = 0.05$. In (E), $n = 3$; in (F), each replicate consisted of six leaves. For NLP2 experiments, $n = 3$ or 4; for NIN experiments, $n = 6$ to 8. In (E) and (F), $^*P < 0.05$; ns, not significant.

non-legumes, followed by recruitment of NLP2, to enhance LgHb expression in legume nodules.

REFERENCES AND NOTES

1. K. Marchal, J. Vanderleyden, *Biol. Fertil. Soils* **30**, 363–373 (2000).
2. L. Schauser, A. Roussis, J. Stiller, J. Stougaard, *Nature* **402**, 191–195 (1999).
3. H. Nishida *et al.*, *Nat. Commun.* **9**, 499 (2018).
4. J. S. Lin *et al.*, *Nat. Plants* **4**, 942–952 (2018).
5. C. W. Liu *et al.*, *Plant Physiol.* **179**, 1704–1722 (2019).
6. M. Konishi, S. Yanagisawa, *Nat. Commun.* **4**, 1617 (2013).
7. F. Berrabah *et al.*, *New Phytol.* **203**, 1305–1314 (2014).
8. T. Hakoyama *et al.*, *Plant Cell Physiol.* **53**, 225–236 (2012).
9. H. Nishida *et al.*, *Plant Cell* **33**, 2340–2359 (2021).
10. T. Soyano, Y. Shimoda, M. Hayashi, *Plant Cell Physiol.* **56**, 368–376 (2015).
11. M. Griesmann *et al.*, *Science* **361**, aat1743 (2018).
12. C. Vázquez-Limón, D. Hoogewijs, S. N. Vinogradov, R. Arredondo-Peter, *Plant Sci.* **191–192**, 71–81 (2012).
13. H. Vigeolas, D. Hühn, P. Geigenberger, *Plant Physiol.* **155**, 1435–1444 (2011).
14. M. Perazzolli *et al.*, *Plant Cell* **16**, 2785–2794 (2004).
15. A. U. Igamberdiev, R. D. Hill, *J. Exp. Bot.* **55**, 2473–2482 (2004).
16. T. Soyano, H. Hirakawa, S. Sato, M. Hayashi, M. Kawaguchi, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14607–14612 (2014).
17. R. C. O'Malley *et al.*, *Cell* **165**, 1280–1292 (2016).
18. L. Zhao *et al.*, *Sci. Rep.* **8**, 1487 (2018).
19. J. M. Alvarez *et al.*, *Nat. Commun.* **11**, 1157 (2020).
20. D. Yan *et al.*, *Nat. Commun.* **7**, 13179 (2016).
21. D. Barker *et al.*, *The Medicago Handbook* (Noble Research Institute, 2006).

ACKNOWLEDGMENTS

We thank X. Fang for providing purified NIN protein and *nlp1* mutant seeds, J. Feng for providing seeds of *nlp2-1*, D. Guan and F. Robson for their contributions early in the project, and G. Oldroyd (University of Cambridge) and the ENSA program for providing plasmid backbones for Golden Gate cloning and *M. truncatula* root transformation. **Funding:** Supported by the CAS Project for Young Scientists in Basic Research (YSBR-011), National Science Foundation China (3217020272), Ministry of Science and Technology (2019FA0904703), Strategic Priority Research Program of the Chinese Academy of Sciences (XDB27040209), National Key R&D Program of China (2016YFA0500500), and Chinese Academy of Sciences (153D31KYSB20160074). Also supported by Biotechnology and Biological Sciences Research Council grants BB/L010305/1 (David Phillips Fellowship), BB/J004553/1 (J.M.), and BB/N003608/1 (C.S.C.); China Postdoctoral Science Foundation grant 2018M642099 (S.J.); ANR grants EPISYM (ANR-15-CE20-0002) and Laboratoire d'Excellence (LABEX) TULIP (ANR-10-LABX-41) (P.G., M.F.J., and Y.P.); and NSF grants DBI-0703285 and IOS-1127155 (K.M. and J.W.). **Author contributions:** The experiments were conducted by S.J., M.-F.J., J.-P.G., Y.P., J.W., F.L., C.S.C., Q.L., Y.T.R., and M.Z.; K.M. and J.W. generated the *Tnt1* insertion mutants; experiments were designed by S.J., M.-F.J., J.-P.G., P.X., P.G., and J.D.M.; and P.X., P.S.P., E.W., P.G., and J.D.M. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg5945
Materials and Methods
Figs. S1 to S14
Tables S1 to S5
References (*22*–*34*)
Data S1
MDAR Reproducibility Checklist

15 January 2021; accepted 16 September 2021
10.1126/science.abg5945

PLANT SCIENCE

Processing of NODULE INCEPTION controls the transition to nitrogen fixation in root nodules

Jian Feng[1], Tak Lee[1,2], Katharina Schiessl[1], Giles E. D. Oldroyd[1,2]*

Legume nodules create an environment for intracellular bacterial symbionts to fix atmospheric nitrogen. The master regulator NODULE INCEPTION (NIN) controls many aspects of nodule initiation, and we demonstrate that it also regulates the transition to nitrogen fixation via proteolytic processing by a signal peptidase complex. Processing of NIN results in a carboxyl-terminal NIN fragment containing the DNA binding motifs, which activates a suite of genes associated with symbiosome development and nitrogen fixation. Similar NIN processing is observed in *Medicago truncatula* and *Lotus japonicus*, implying a conserved mechanism of cell state transition. These findings explain how legume nodules transition to a nitrogen-fixing state and a mechanism by which a single transcription factor can regulate many different developmental processes necessary in the activation and regulation of nitrogen fixation.

Some angiosperms, including legumes, associate with nitrogen-fixing bacteria within nodules, which are modified lateral roots (*1*, *2*). Nodules create the low-oxygen environment for bacterial nitrogenase to convert dinitrogen to ammonium. Nodule initiation and promotion of intracellular bacterial infection is facilitated by *NODULE INCEPTION* (*NIN*) (*3*, *4*), in part through transcriptional activation of *NUCLEAR FACTOR-Y SUBUNIT A1* (*NF-YA1*), *NODULE PECTATE LYASE* (*NPL*), *CYTOKININ RESPONSE 1* (*CRE1*), and *LATERAL ORGAN BOUNDARIES DOMAIN 16* (*LBD16*) (*5*, *6*). *NIN* also controls nodule number through mobile CLE peptides that promote root-shoot-root signaling (*7*, *8*). *NIN* is one of only a few genes that are consistently lost with losses of nodulation over evolutionary time (*9*, *10*). Here we demonstrate that NIN also controls maturation of the nodule to the nitrogen-fixation state through proteolytic processing by a nodule-specific signal peptidase complex (SPC).

A series of functional NIN–green fluorescent protein (GFP) fusions spanning the full length of the protein (table S1) all showed nuclear localization (Fig. 1E), consistent with multiple predicted nuclear localization signals (Fig. 1A). An array of antibodies produced against NIN peptides (Fig. 1A) revealed a predicted full-length NIN and a second larger band, possibly an oligomeric form, when *NIN* in transiently expressed in *Nicotiana benthamiana* (Fig. 1B and fig. S1). The same antibodies showed no NIN-specific bands in total root and nodule preparations (fig. S2) but did show clear NIN-specific bands from nuclei preparations (Fig. 1C and figs. S2 and S3). However, the predicted full-length product was only faintly observed with α-RWP in wild type (Fig. 1C and fig. S3A), as well as with α-internal region (IR) in hypernodulating *sunn* mutant (*11*) (fig. S4). In place of the full-length product, we observed two smaller products: a c40kDa product detected by α-N terminus (NT) and a c55kDa product with α-IR, α-RWP, and α-C terminus (CT) (Fig. 1C). α-RWP also revealed additional products not detected by the other antibodies (Fig. 1C). All of these bands were absent in the *nin-4* (Fig. 1C) null mutant (fig. S5). The c55kDa product must possess the DNA binding domain (RWP-RK) and the protein-protein interaction domain (PB1), because antibodies that detect this product span these domains (Fig. 1A). To validate these observations, we generated functional (fig. S6) 3×FLAG NIN-fusions: α-FLAG antibodies detected the c55kDa C-terminal and predicted full-length products (Fig. 1D) but not the c40kDa product.

The processing of NIN is not the result of alternative splicing of *NIN* mRNA and is in-

[1]Sainsbury Laboratory, University of Cambridge, 47 Bateman Street, Cambridge CB2 1LR, UK. [2]Crop Science Centre, University of Cambridge, 93 Lawrence Weaver Road, Cambridge CB3 0LE, UK.
*Corresponding author. Email: gedo2@cam.ac.uk

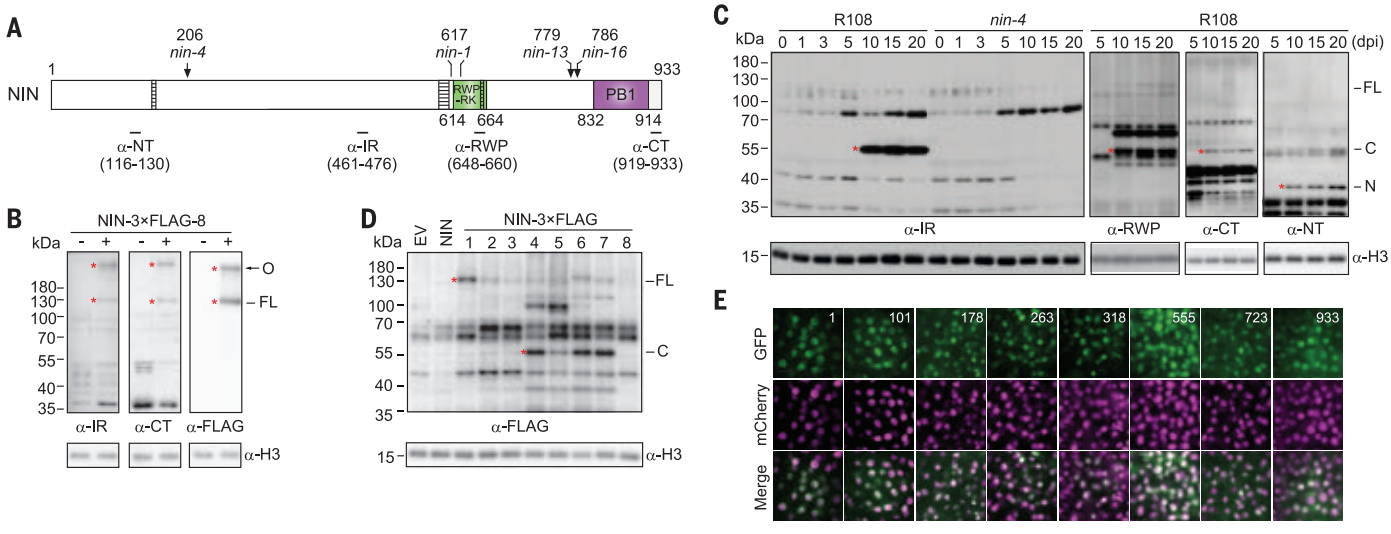


Fig. 1. NIN is processed during nodule formation. (**A**) A schematic of NIN showing the RWP-RK and PB1 domains and predicted nuclear localization signals (hatching). The position of peptides used for antibody generation and mutant alleles are indicated. (**B**) Immunoblot analysis of NIN-3×FLAG proteins expressed in *N. benthamiana* leaves using α-IR, α-CT, and α-FLAG antibodies. (**C**) Immunoblot analysis with α-NIN in the nuclear extracts of wild-type (R108) and *nin-4* roots at indicated time points after *S. meliloti* inoculation. dpi, days post inoculation. (**D**) Different NIN-3×FLAG fusions expressed in *M. truncatula* (R108) roots and detected with α-FLAG antibodies. EV, empty vector; NIN, full-length *NIN* and *NIN-3×FLAG* fusions (numbers refer to different positions of 3×FLAG) (fig. S6). (**E**) Nuclear localization of NIN-GFP fusions expressed in nodules. Nuclei are labeled with 3×mCherry fused to a nuclear localization signal. Numbers refer to different positions of the GFP tag, as shown in table S1. Red asterisks in (B) to (D) indicate NIN-specific products. For (B) to (D), equal loading was verified with α-histone H3. O, oligomeric form of NIN; FL, predicted full-length NIN; C, c55kDa C-terminal product; N, c40kDa N-terminal product.

sensitive to cycloheximide and MG132, inhibitors of protein synthesis and the proteasome (fig. S7). We conclude that NIN appears to be proteolytically processed at a site internal to the protein to produce a c40kDa N-terminal product and a c55kDa C-terminal product, which are only apparent 10 days after *Sinorhizobium meliloti* infection (Fig. 1C and fig. S8). *DNF1* is a nodulation-induced component of a SPC (fig. S9) that is expressed in the zone of the nodule where bacterial infection occurs (*12*). The DNF1 complex has been implicated in processing the signal peptide from nodule-specific cysteine-rich (NCR) peptides (*12*, *13*), which are necessary for bacteroid development (*14*, *15*). *dnf1* mutants produce small, white nodules (*12*) similar to two weak *nin* alleles (fig. S5, G to I) (*16*, *17*), and all show sparse intracellular bacteria that lack differentiation (Fig. 2B). These internalized bacteria do not transition into the nitrogen-fixing state [nitrogenase iron protein (NifH) is lacking] (Fig. 2A), and the nodules lack expression of genes such as leghemoglobins and many NCR and glycine-rich peptides (GRPs) (Fig. 3, A to C).

The similarity in the mutant phenotypes of *dnf1-2* and *nin-16* led us to assess whether the DNF1 complex may be associated with the proteolytic processing of NIN. To this end, we assessed NIN accumulation in *dnf1-2* and found neither the c40kDa nor the c55kDa products (Fig. 2C), which cannot be explained by the effects of *dnf1* mutation on transcription of *NIN* (Fig. 3B and fig. S10). The catalytic subunit of SPC is conserved in eukaryotes (fig. S11A), and, on the basis of prior knowledge (*18*), we predicted a NIN cleavage site at A400 (fig. S11, B and C, and table S5). Mutation of A400 blocked processing (fig. S11D) and recapitulated the *dnf1-2* and *nin-16* phenotypes (Fig. 2D). This alanine residue is conserved in NIN-like proteins (NLPs), but the neighboring structure differs, which may affect the ability of SPC to process NLPs (fig. S11, E and F). Unexpectedly, we did not observe reciprocal accumulation of full-length NIN in any of the mutants that affect processing; only a small accumulation of full-length NIN was observed after treatment with MG132 in *dnf1-2* (fig. S12). We propose that full-length NIN is unstable, and degradation of full-length NIN may be further facilitated by other induced proteases in *dnf1-2* (fig. S13).

The SPC is located on the endoplasmic reticulum membrane (*19*), and we found that NIN c40kDa and c55kDa were associated with membrane and nuclear fractions, with nuclear preference (Fig. 2C and fig. S14). SPC primarily removes signal peptides of proteins targeted for secretion (*20*), however, other unusual targets, such as NIN, have been identified, and the human SPC processes viral proteins lacking signal peptides (*21*, *22*).

Overexpression of *NIN* in the absence of *S. meliloti* leads to the accumulation of low

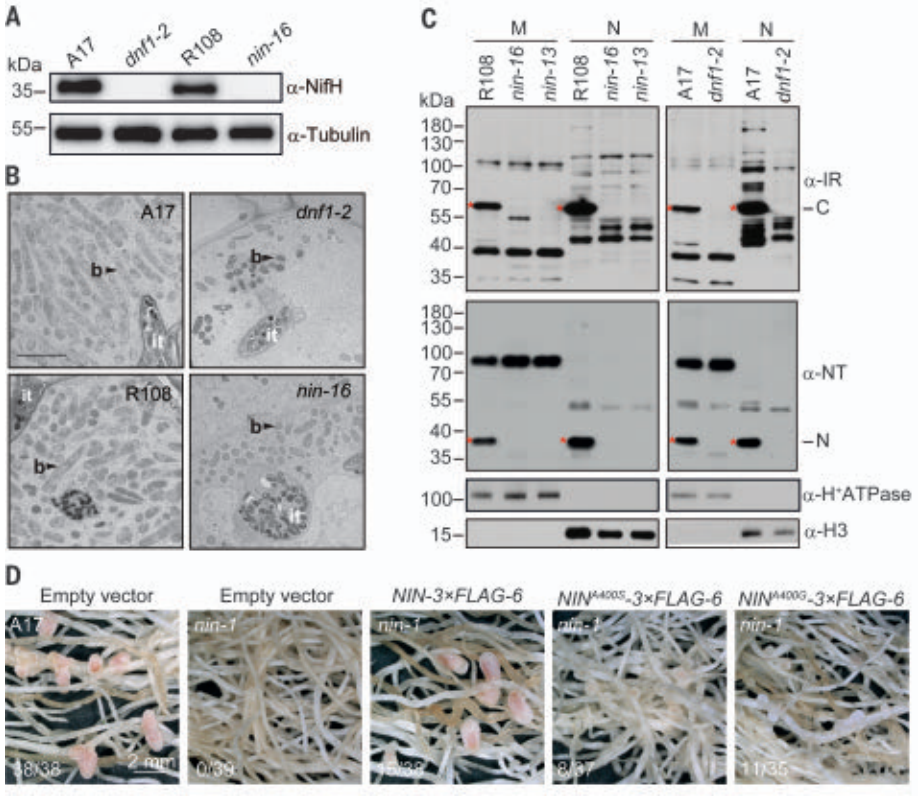


Fig. 2. DNF1 complex–mediated processing of NIN during the transition to nitrogen fixation. (**A**) Bacterial NifH in the nodules of wild-type (A17 and R108), *dnf1-2*, and *nin-16*. (**B**) Transmission electron microscopy (TEM) images showing infection threads (it) and bacteroids (b) in wild-type (A17 and R108), *dnf1-2*, and *nin-16* nodules. Scale bar, 2 μm. (**C**) The c40kDa and c55kDa processed NIN products are present in membrane (M) and nuclear fractions (N) of wild-type (A17 and R108) but absent in *nin-13*, *nin-16*, and *dnf1-2*. Red asterisks indicate NIN-specific products. α-H⁺ATPase (hydrogen adenosine triphosphatase) is loading control for M, and α-histone H3 is the control for N. (**D**) Wild-type (A17) and *nin-1* transformed with *NIN* and *NIN* A400 mutants, 28 days after *S. meliloti* inoculation. Number of transformed plants with nodules relative to total number of transformed plants is shown for each image.

levels of NIN c55kDa and higher levels of the predicted full-length product (fig. S15). We suggest that processing of overexpressed NIN results from the constitutively expressed SPC (fig. S9B). Overexpression of *NIN* is sufficient to activate nodulation (fig. S16) and the expression of many late genes associated with nitrogen fixation that were not expressed in *nin-16* and *dnf1-2* (Fig. 3, A to C). Transactivation assays reveal that the processed C-terminal form of NIN can promote the expression of late nodule-associated genes but not early nodule-associated genes (Fig. 3D and fig. S17) from which we conclude that NIN is sufficient to activate early genes involved in nodule organogenesis and late genes associated with nitrogen fixation, but the latter step requires processing by the DNF1 SPC.

The DNF1 SPC was proposed to process NCR peptides (*12*, *13*), but here we suggest an additional, or alternative, target. Whereas *NIN* is essential in many legumes, including *Lotus japonicus*, NCR peptides are limited to legume species that show bacteroid differentiation (*14*, *23*), which excludes *L. japonicus*. To assess the relative importance of the potential targets of DNF1 processing, we generated a *dnf1* mutant in *L. japonicus* (*Ljdnf1*) (fig. S18). This mutant recapitulated the phenotypes observed in *M. truncatula dnf1*: small, white nodules with few bacteria and reduced accumulation of NifH (Fig. 4, A and B). In addition, while symbiosomes in wild-type nodules often contained multiple bacteria, symbiosomes of *Ljdnf1* contained only single bacteria, whose membranes were deformed, providing evidence of premature degradation (Fig. 4C). Using antibodies generated against *L. japonicus* NIN (α-LjNIN), we found evidence for NIN processing in *L. japonicus*, comparable to that observed in *M. truncatula* (Fig. 4, D and E, and fig. S19), with c60kDa and c38kDa fragments as primary products (Fig. 4, D and E). As in *M. truncatula*, accumulation of the processed forms occurred 10 days after inoculation with *Mesorhizobium loti* and were depleted in *Ljdnf1* (Fig. 4E). The combination of comparable phenotypes for *dnf1* mutants, along with the comparable processing

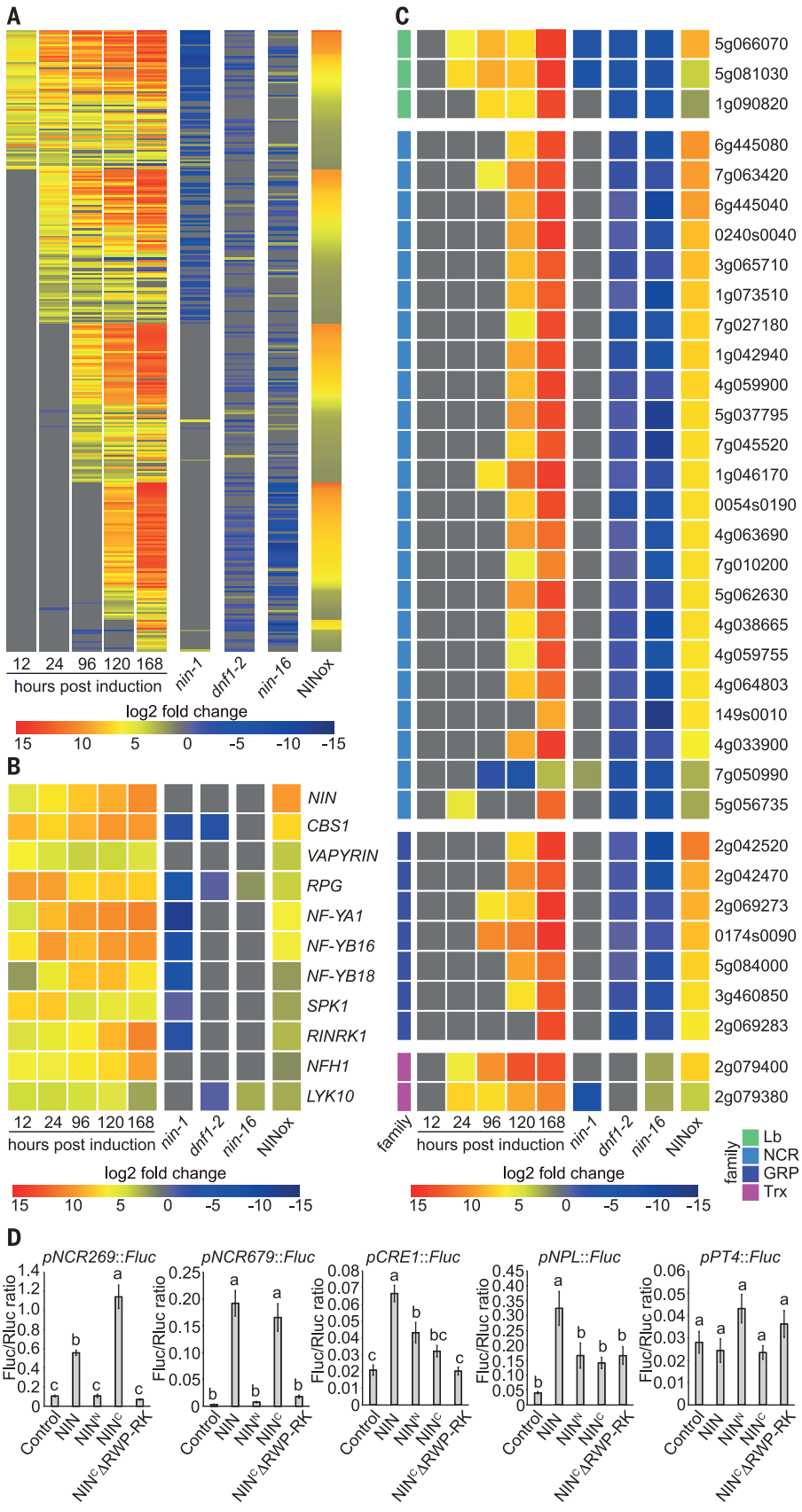


Fig. 3NIN is necessary and sufficient for early and late nodulation-associated gene expression. (**A**) Genes up-regulated by *NIN* overexpression (NINox) that are also induced after treatment with *S. meliloti*, shown in wild type and *nin-1* mutants 12 to 168 hours and 24 hours, respectively, after *S. meliloti* inoculation and in *dnf1-2* and *nin-16* nodules (log2 fold changes for the mutants are relative to the wild type with the equivalent treatment). (**B** and **C**) Heatmaps of selected genes from (A). Lb, leghemoglobin; Trx, thioredoxin. (**D**) Transactivation assays. Fluc, firefly luciferase; Rluc, *Renilla* luciferase. Data are mean ± SEM (n = 6 biological replicates). Letters a, b, and c denote statistically significant groupings analyzed by one-way analysis of variance with Tukey's test, $P < 0.05$.

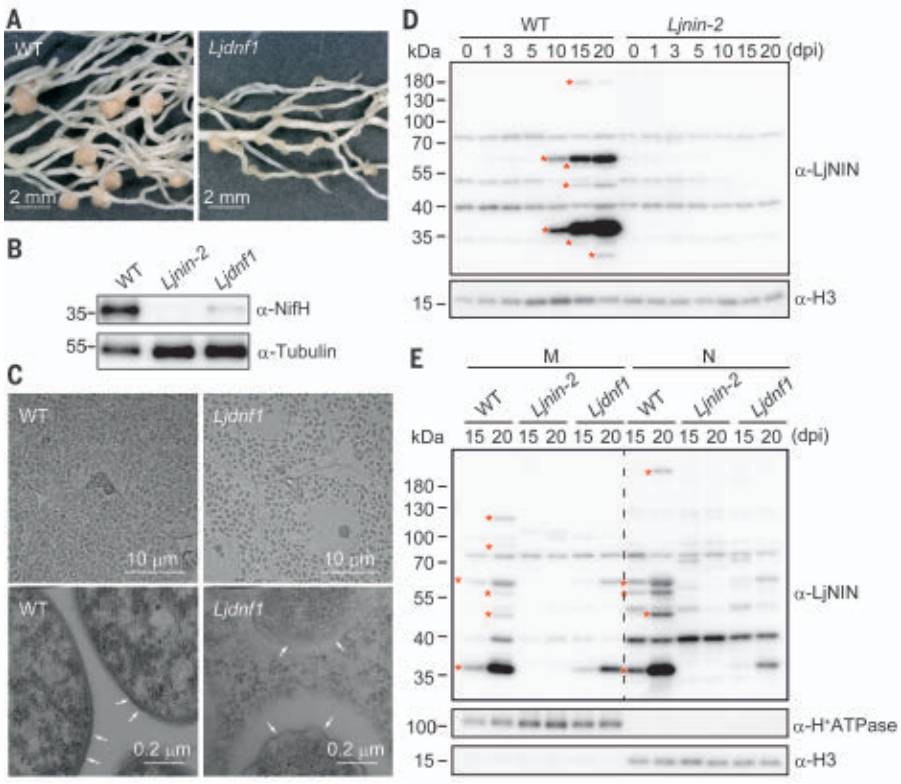


Fig. 4. DNF1-dependent NIN processing is conserved in *L. japonicus*. (**A**) Wild-type (WT) and *Ljdnf1* nodules 14 days after *M. loti* inoculation. (**B**) Bacterial NifH in the nodules of WT, *Ljnin-2*, and *Ljdnf1*. (**C**) TEM images showing symbiosomes in nodules of WT and *Ljdnf1*. Arrows indicate bacterial membranes. (**D**) LjNIN proteins in the nuclear extracts of WT and *Ljnin-2* at the indicated time points after inoculation of *M. loti*. (**E**) LjNIN in the membrane fractions (M) and nuclear fractions (N) of WT, *Ljnin-2*, and *Ljdnf1*. Red asterisks in (D) and (E) indicate NIN-specific products.

of NIN in both species, leads us to conclude that NIN is a target for the DNF1 complex in both species and that NCR processing alone cannot explain the *dnf1* mutant phenotypes.

NIN controls all aspects of symbiotic nodulation: initiation of nodule organogenesis, intracellular rhizobial infection, regulation of nodule number (*6*), and now also the transition to nitrogen fixation. We show that one mechanism allowing this single protein to control so many different functions is proteolytic processing by the DNF1 SPC. The specific expression of *DNF1* in the infection zone of the nodule (*12*) likely allows transition of cells into those that can support nitrogen fixation and sustain bacterial infection. While *NIN* is essential for early and late nodule-associated gene expression, *DNF1* is essential only for the late nodule-associated gene expression. Thus, processing of NIN is only relevant for functions late in nodule development, not for activation of earlier genes such as *CRE1* and *NPL*. Among the genes controlled by *NIN* at late stages of nodulation are leghemoglobins that buffer oxygen (*24*), NCR peptides that drive bacteroid differentiation (*15*), and thioredoxins that control the redox state of the nodule (*25*). NIN, as processed by the DNF1 complex, is both necessary and sufficient for the expression of these genes and the transition to the nitrogen-fixing state.

REFERENCES AND NOTES

1. K. Schiessl *et al.*, *Curr. Biol.* **29**, 3657–3668.e5 (2019).
2. T. Soyano, Y. Shimoda, M. Kawaguchi, M. Hayashi, *Science* **366**, 1021–1023 (2019).
3. L. Schauser, A. Roussis, J. Stiller, J. Stougaard, *Nature* **402**, 191–195 (1999).
4. J. F. Marsh *et al.*, *Plant Physiol.* **144**, 324–335 (2007).
5. R. Shahan, P. N. Benfey, *Dev. Cell* **52**, 6–7 (2020).
6. S. Roy *et al.*, *Plant Cell* **32**, 15–41 (2020).
7. T. Soyano, H. Hirakawa, S. Sato, M. Hayashi, M. Kawaguchi, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14607–14612 (2014).
8. D. Tsikou *et al.*, *Science* **362**, 233–236 (2018).
9. R. van Velzen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4700–E4709 (2018).
10. M. Griesmann *et al.*, *Science* **361**, eaat1743 (2018).
11. E. Schnabel, E. P. Journet, F. de Carvalho-Niebel, G. Duc, J. Frugoli, *Plant Mol. Biol.* **58**, 809–822 (2005).
12. D. Wang *et al.*, *Science* **327**, 1126–1129 (2010).
13. W. Van de Velde *et al.*, *Science* **327**, 1122–1126 (2010).
14. E. Kondorosi, P. Mergaert, A. Kereszt, *Annu. Rev. Microbiol.* **67**, 611–628 (2013).
15. G. Maróti, J. A. Downie, É. Kondorosi, *Curr. Opin. Plant Biol.* **26**, 57–63 (2015).
16. V. Veerappan *et al.*, *BMC Genomics* **17**, 141 (2016).
17. J. Liu *et al.*, *New Phytol.* **230**, 290–303 (2021).
18. N. D. Rawlings *et al.*, *Nucleic Acids Res.* **46**, D624–D632 (2018).
19. H. A. Meyer, E. Hartmann, *J. Biol. Chem.* **272**, 13159–13164 (1997).
20. M. Paetzel, A. Karla, N. C. Strynadka, R. E. Dalbey, *Chem. Rev.* **102**, 4549–4580 (2002).
21. D. Estoppey *et al.*, *Cell Rep.* **19**, 451–460 (2017).
22. R. Zhang *et al.*, *Nature* **535**, 164–168 (2016).
23. J. Montiel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5041–5046 (2017).
24. T. Ott *et al.*, *Curr. Biol.* **15**, 531–535 (2005).
25. C. W. Ribeiro *et al.*, *Curr. Biol.* **27**, 250–256 (2017).

ACKNOWLEDGMENTS

We thank J. Murray, R. Dickstein, J. Stougaard, and J. Fournier for providing strains and seeds and E. Barclay and K. Findlay for support with microscopy. **Funding:** This work was supported by the Bill and Melinda Gates Foundation and the UK Foreign, Commonwealth and Development Office (OPP1028264) through Engineering the Nitrogen Symbiosis for Africa (ENSA) project; the Biotechnology and Biological Sciences Research Council (BB/K003712/1); and the Gatsby Foundation (GAT3395/GLH).
Author contributions: J.F. and G.E.D.O. designed the experiments, analyzed the data, and wrote the manuscript. J.F. performed most of the experiments. K.S. performed *NIN* overexpression. T.L. analyzed RNA sequencing (RNA-seq) data.
Competing interests: The authors declare no competing interests.
Data and materials availability: All data are available in the main text or the supplementary materials. The RNA-seq data generated in this study have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE178119.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg2804
Materials and Methods
Figs. S1 to S19
Tables S1 to S6
References (*26*–*59*)
MDAR Reproducibility Checklist

11 January 2021; accepted 27 August 2021
10.1126/science.abg2804



EMPOWER WITH EVIDENCE

AAAS | ANNUAL MEETING

See inside for details:

Hybrid Format Information | Plenary Sessions | Topical Lectures
Meeting Highlights | Registration Information

The hybrid 2022 AAAS Annual Meeting will be held in Philadelphia, and online, February 17-20. Registration and program are now available!

For more information, please visit:

aaas.org/meetings | #AAASmtg



Dear Science Enthusiast,

After a year of attending virtual convenings, we've come to appreciate many benefits, but we still long for the things that can only happen in person. This year's meeting theme **"Empower With Evidence"** aims to inspire sessions that illustrate the transformative power of evidence-based discoveries and that address the challenges of rebuilding the public trust in science.

As vaccinations increase, case numbers decline, and states, localities, and institutions begin the process of reopening, we are all left wondering what the newest "new normal" will look like. Like most scientific organizations, we are actively confronting these questions, especially as they relate to our 2022 Annual Meeting:

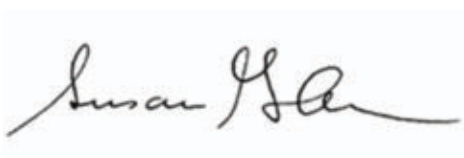
- **Philly, here we come!** We invite those who are comfortable doing so to join us February 17-20 in Philadelphia. Our in-person experience will feature live Plenary Lectures, Topical Lectures, special programming, and great social and networking opportunities—all in a city that is ready to welcome visitors—safely—with open arms.

- Effective September 1, 2021, AAAS will require that **all participants in its in-person meetings and events**—including, but not limited to, presenters, attendees, exhibitors, staff, and guests—**be fully vaccinated.**

- **A first-rate virtual experience**. Not able to join us in Philly? No problem! You'll be able to participate in all of our Plenary Lectures, Topical Lectures, Scientific Sessions, and Workshops—live!—via a new and improved virtual meeting platform.

The Annual Meeting team will continue to work tirelessly to create a convening that features compelling content, powerful professional development, and unparalleled networking opportunities in the safest possible settings. We look forward to seeing you in Philly or online.

Sincerely,
Susan G. Amara, Ph.D.
President, AAAS
Chair, 2022 Annual Meeting



Join us in Philadelphia or Online for the 2022 AAAS Annual Meeting

AAAS requires that all participants in its in-person meetings be fully vaccinated.

Connect with STEM professionals, policymakers, funders, educators, students, and industry leaders, as you participate in the hybrid **#AAASmtg** programming:

Deluxe Access registrants can:

- Join **Scientific Sessions** online or in Philadelphia, and have access to the pre-released library of **Spotlight Video**
- Be there for **lectures from renowned speakers** in academia and industry
- Take advantage of opportunities to **meet new collaborators** and reconnect with old friends
- Participate virtually with features included in the Basic Access registration

Basic Access registrants can:

- **Attend workshops** and learn new ways to forward their research trajectory
- Wander the virtual halls of the **E-poster Session** and talk to finalists about their work
- Tune into **STEM documentaries**, **podcasts**, and other **special sessions**.

Find the full listing of the 2022 AAAS Annual Meeting program, including scientific session, workshops, posters and more at **aaas.org/meeting**

CONNECT WITH US!
#AAASmtg

aaas.org/meetings

@AAASmeetings

/AAAS.Science



Plenary Sessions



AAAS President Address

Susan Amara
President, AAAS
Chair, 2022 Annual Meeting



Addressing Energy Access, Equity and Resiliency in the 21st Century

Deepakraj M. Divan
Director, Center for Distributed Energy, Georgia Tech



Katalin Karikó
Senior Vice President, BioNTech RNA Pharmaceuticals
Adjunct Professor, University of Pennsylvania School of Medicine



"Does Science Communication Still Work?"

A debate moderated by **Holden Thorp**

Editor-in-Chief of Science Journals

Topical Lectures



Bioinspired Ocean Exploration

John O. Dabiri

California Institute of Technology



From the Possibility to the Certainty of a Supermassive Black Hole

Andrea M. Ghez

University of California, Los Angeles



Games, Stories, and Science for Successful Public Engagement

Katie Hinde

Arizona State University



The Disappearing Chip: The Physics behind Nanoscale Electronics

Nadya Mason

University of Illinois at Urbana-Champaign



A Care Conundrum at the Intersection of Class, Access, and Health

Robin Nelson

Arizona State University



Unraveling the Big Lie: Participatory Disinformation and its Threat to Democracy

Kate Starbird

Human Centered Design and Engineering



Against Crisis Science: Research Futures for Climate and Energy Justice

Kyle Whyte

University of Michigan

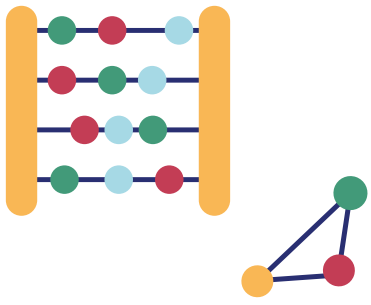


Evidentiary Constraint: When Science Policy Meets Activism

SARTON MEMORIAL LECTURE IN THE HISTORY AND PHILOSOPHY OF SCIENCE

Audra J. Wolfe

Writer and Editor



As of October 13, 2021

Join us in Philadelphia or Online for the 2022 AAAS Annual Meeting

AAAS requires that all participants in its in-person meetings be fully vaccinated.

The 2022 AAAS Annual Meeting will Empower With Evidence by featuring groundbreaking multi-disciplinary research—not only to advance knowledge, but to understand, and, ultimately, resolve many of the problems we face today.

BE A PART OF THE CONVERSATION

Public Health | Science Communication | Big Data | Artificial Intelligence
Stem Education | Science Policy | Social and Behavorial Sciences | Innovation
Diversity and Equity | Climate and Sustainibilty | Linguistics | Global Perspectives

and more...

CONNECT WITH US! #AAASmtg

aaas.org/meetings

@AAASmeetings

/AAAS.Science

ADVANCE DELUXE REGISTRATION RATES	AAAS Member Deluxe Access	Non-Member Deluxe Access	Delux Access: Virtual Only
General Attendee	$355	$500	$240
Postdoc	$155	$300	$155
Retired Professional	$285	$415	$240
Student	$75	$110	$75

Advance registration is available now through **January 24, 2022**.
On-site registration rates will apply thereafter.

Get in touch with us about sponsorship and exhibit opportunities in Philadelphia and meeting virtual platform: meetings.aaas.org/sponsors-exhibitors.



Ninth International Congress on Peer Review and Scientific Publication

Sept 8-10, 2022 | Swissôtel | Chicago, IL

Congress Advisory Board

Congress Directors
John P.A. Ioannidis, Meta-Research Innovation Center at Stanford (METRICS)

Michael Berkwits, *JAMA* and the JAMA Network

Congress Executive Director
Annette Flanagin, *JAMA* and the JAMA Network

European Director
Fiona Godlee, *BMJ*

European Coordinator
Theodora Bloom, *BMJ*

Associate Directors
Steve Goodman, Meta-Research Innovation Center at Stanford (METRICS)

Véronique Kiermer, *PLOS*

David Moher, Ottawa Hospital Research Institute

Director Emeritus
Drummond Rennie, University of California, San Francisco

Vivienne Bachelet, *Medwave,* Universidad de Santiago de Chile

Albert-László Barabási, Northeastern University

Howard Bauchner, Boston University School of Medicine

Lisa A. Bero, The University of Sydney

Patrick M. Bossuyt, University of Amsterdam

Lex Bouter, Vrije University

Isabelle Boutron, Université de Paris

An-Wen Chan, University of Toronto

Kay Dickersin, Johns Hopkins Bloomberg School of Public Health

Dan Evanko, American Association for Cancer Research

James Evans, The University of Chicago

R. Brian Haynes, McMaster University Faculty of Health Sciences

James Kigera, *Annals of African Surgery*

Sabine Kleinert, *The Lancet*

Christine Laine, *Annals of Internal Medicine*

José Florencio F. Lapeña Jr, University of the Philippines

Malcolm MacLeod, The University of Edinburgh

Emilie Marcus, University of California, Los Angeles

Ana Marušić, *Journal of Global Health*

Bahar Memani, Elsevier

Frank Miedema, University Medical Center Utrecht

Jigisha Patel, London, UK

Eric J. Rubin, *New England Journal of Medicine*

David Schriger, University of California, Los Angeles

Emily Sena, *BMJ Open Science*

Magdalena Skipper, *Nature* and Nature Research

Victoria Stodden, University of Illinois

Deborah Sweet, Cell Press

Sarah Tegen, American Chemical Society

Martin R. Tramèr, Geneva University Hospitals

Valda Vinson, *Science*

Erik Von Elm, Cochrane Switzerland

Eric-Jan Wagenmakers, University of Amsterdam

Jasmine Wallace, American Society for Microbiology

Rebecca Williams, National Library of Medicine

CALL FOR ABSTRACTS

Abstracts are due January 31, 2022

***JAMA, BMJ*, and METRICS invite abstracts for the Ninth International Congress on Peer Review and Scientific Publication.**

The Congress will provide a forum for the presentation and discussion of new research into the quality and credibility of peer review and scientific publication, to establish the evidence base on which scientists can improve the conduct, reporting, and dissemination of scientific research.

Abstracts reporting new research on any aspect of editorial and funding peer review, scientific publication, research and reporting practices, and information exchange will be considered.

Topics of Interest, include

- Bias
- Editorial and Peer Review Decision Making
- Research and Publication Ethics
- Improving the Quality of Reporting
- Models for Peer Review and Scientific Publication
- Dissemination of Scientific and Scholarly Information

More details, examples of Topics of Interest, Instructions for Preparing and Submitting Abstracts, and abstracts of previously presented research are available at **peerreviewcongress.org**

Abstracts may be submitted December 1, 2021 through January 31, 2022

Questions? jama-peer@jamanetwork.org

JAMA Network™ thebmj METRICS META-RESEARCH INNOVATION CENTER AT STANFORD

peerreviewcongress.org



Science 2021
TOP EMPLOYERS

2021's Top Employers engage, nurture, and harness the power of scientists

While respondents in the Top Employers Survey say they are looking to work for companies that are innovative, another characteristic they look for is a close second: respect. **By Alaina G. Levine**

Ask Christina Nurse, the Ph.D. senior biostatistician manager at **Takeda Pharmaceutical Company Ltd.**, why she loves her employer. She's clear and confident: "I get to work with some of the most brilliant scientific minds who care so deeply about bringing medicines to our patients." But a personal story hints at something more. Just as she returned from maternity leave in March 2020, as the pandemic was proliferating across the globe, Nurse's new son needed heart surgery. Takeda's senior leadership called her to ask if she needed anything and how they could help. "It reminded me that I had the flexibility to care for my family," she says, "and it's because of that care and compassion that I can be my best." Her son is doing "amazing," and so is Nurse, because she has an employer that sees and values her as a person.

Too often, companies say they care as part of public relations campaigns. But amazing companies are those who invest real resources in supporting the holistic growth of their most vital asset—the staff—and step up in difficult, unpredictable, and even scary circumstances. There are people behind these top companies—and it is the people who make them top employers. "People say companies don't have a soul, but there is a soul at Takeda. It's not just posters on walls," says Andy Plump, president of R&D at Takeda (ranked No. 14), which is based in Tokyo, Japan.

Takeda and many other biotech and pharmaceutical firms are recognized as this year's top employers by the journal *Science*. Now in its 20th year, *Science*'s Top Employers Survey is meant to celebrate and highlight the best of the best life sciences companies from the employee's perspective. Earlier this year, *Science* conducted an online survey of readers and received 7,225 responses, with an almost 30% split between those with Ph.D.s, Master's, and Bachelor's degrees, about 51% of whom work in research or development.

The objectives of the survey are to reveal the characteristics that most drive biotech and pharma organizations' reputations and to flesh out what makes them so great. This year's survey focuses on changes the companies have experienced as a result of the pandemic, the resulting adjustments they made to their culture and ecosystem, and the opportunities they provide for employees to soar. It also gives us a peek into how these organizations pivoted to support COVID-19 testing and vaccine development and rollouts.

Putting staff and patients at the center

A word cloud, assembled from comments made by respondents, offers the first clue of what makes a top employer: In the center and in the largest-size font is the word "Employee." It is not surprising that the most appreciated companies are those that center the staff experience in their cultures, and when we engage leadership about why their company was selected, almost all of the top companies say it is because they are employee-focused. "At Incyte, our roles have high visibility and high accountability, and respect for employees is of the highest priority," says Paula Swain, executive vice president of human resources at **Incyte**, a Wilmington, Delaware–based biotech company (No. 2). "You can have a real impact here. We have a relatively flat organization, so there is a broad understanding of what is going on throughout the company, and everyone has the ability to interact with leadership and across functional teams. The door is literally open all the time."

But employees also rank top companies by their dedication to patients. "Our culture is something our employees are exceptionally proud of," says Nicole Schaeffer, chief

cont.>

PHOTO: © ANDREY_POPOV/SHUTTERSTOCK.COM

POSTDOCTORAL TRAINING PROGRAM

PURSUE YOUR PASSION FOR SCIENCE IN A DYNAMIC AND INNOVATIVE ENVIRONMENT



Applications Accepted
October 1 to December 1, 2021
Start dates flexible, but if possible,
Summer 2022

- Up to four years of formal postdoctoral training in a program co-Directed by **award-winning and published educators,** both of whom are working scientists at Regeneron **devoted to training postdoctoral fellows**
- Exposure to the biotechnology industry's **most successful and innovative scientists**
- **Weekly postdoctoral scientist meetings** to foster scientific rigor, creative thinking, critical reasoning, and career advancement
- A **prestigious, well resourced, and dynamic** industry environment
- Encouragement of **scholarly productivity** such as publications and conference presentations
- Opportunities to pursue your own **cutting edge research ideas** in almost any area of biomedicine, including human genetics

LEARN MORE AT
www.regeneron.com/scienceeducation_professional

REGENERON
SCIENCE TO MEDICINE®

Regeneron is an equal opportunity employer. As such,applicants are considered for employment without regard to race, color, religion, sex, national origin, age, sexual orientation, marital status, disability, genetic information, military or veteran status, or any other category protected under applicable federal, state, or local law.

Top Twenty Employers



2021 Rank	2020 Rank	Employer (global headquarters)	Innovative leader in the industry	Treats employees with respect	Is socially responsible	Has loyal employees	Easily adapts to change
1	–	Insmed (Bridgewater, NJ)	•	•	•		
2	2	Incyte (Wilmington, DE)	•	•	•		
3	3	Alnylam Pharmaceuticals (Cambridge, MA)	•	•	•		
4	1	Regeneron (Tarrytown, NY)	•		•	•	
5	10	Spark Therapeutics (Philadelphia, PA)	•	•	•		
6	4	Syngenta Group (Basel, Switzerland)	•	•	•		
7	9	Moderna (Cambridge, MA)	•	•			•
8	8	Vertex Pharmaceuticals (Boston, MA)	•	•	•		
9	6	Merck KGaA (Darmstadt, Germany)		•	•	•	
10	11	Genentech (South San Francisco, CA)	•	•	•		
11	14	Abbott (Abbott Park, IL)		•	•	•	
12	5	Biocon Limited (Bangalore, India)	•		•	•	
13	12	AbbVie (North Chicago, IL)		•	•	•	
14	18	Takeda Pharmaceutical Company Limited (Tokyo, Japan)		•	•	•	
15	7	Novo Nordisk (Bagsværd, Denmark)	•		•	•	
16	17	Pfizer (New York, NY)	•	•	•		
17	13	GlaxoSmithKline (GSK) (London, United Kingdom)		•	•	•	
18	16	Eli Lilly and Company (Indianapolis, IN)		•	•	•	
19	19	Novartis (Basel, Switzerland)	•	•	•		
20	15	Roche – excluding Genentech (Basel, Switzerland)	•		•	•	

The 20 companies with the best reputations as employers and the top three driving characteristics for each company, according to respondents in the 2021 survey undertaken for the *Science*/AAAS Custom Publishing Office.
The company without a 2020 rank did not receive enough mentions to qualify or did not receive a high enough ranking from the 2020 survey.

people strategy officer at Bridgewater, New Jersey–based **Insmed** (No. 1). "We are a patient-centered, patient-driven culture." Walter Perkins, chief technology officer at Insmed, explains the origin of this tenet: "Our CEO said put the patient first, use that as your compass, and everything will take care of itself in the business." To bring home this point, Insmed, like several top employers, regularly invites patients to speak directly with the employees, an effort that empowers its staff to more actively engage in furthering the company's mission of developing life-saving interventions with patient outcomes in mind. "It is hard to understand how devastating these diseases are, so to hear [patients'] intimate stories of what they go through on a daily basis is striking," says Perkins.

The Top Employers Survey results reveal additional fascinating finds: "Organizations that are seen as innovative leaders in the industry, who easily adapt to change, who respect their employees, have loyal employees, and are socially responsible, are selected as best in class." In fact, being an 'innovative leader' continues to be the most powerful statistical driver of the selection of best companies. Interestingly, this is the first time in 20 years that the adaptability of a company is listed as paramount to the employee experience, and one can guess that this is related to employers' responses to the pandemic and COVID-19.

COVID-19 response

COVID-19 has been a nightmare for humanity, but it has also brought out the very best of human ingenuity. Top employers harnessed the diverse talents of a global human resources pipeline and infrastructure to mobilize strategic responses.

"We surprise people with how agile we are," says Vildan Kehr, divisional vice president of human resources for the rapid and molecular diagnostics division of **Abbott** (No. 11), a multinational medical device and health care company with headquarters in Abbott Park, Illinois. "With COVID, we saw the calling. We had to do something. In 3 months, we built a factory to produce COVID tests. There's nothing like this in history. But the nation and people needed us."

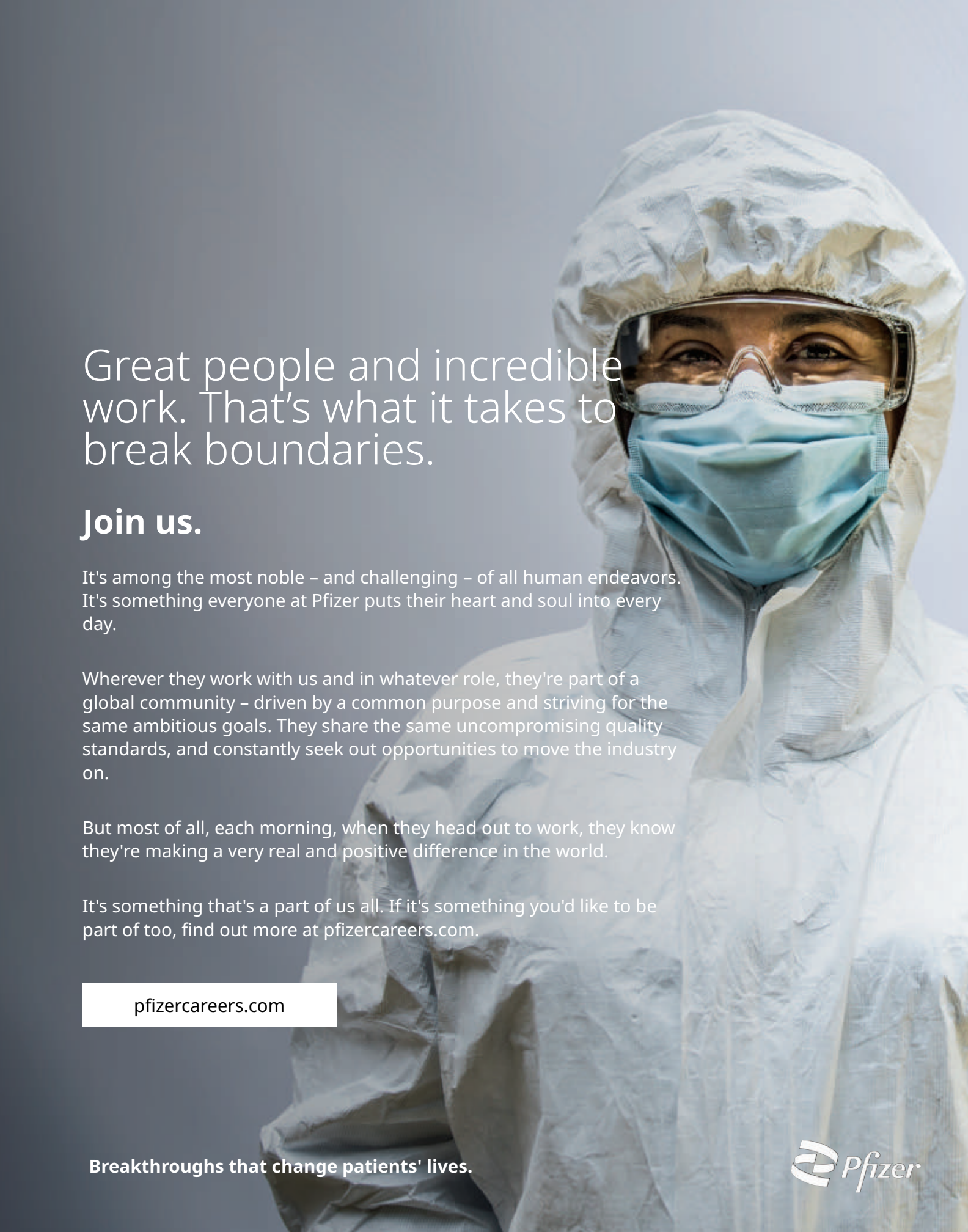
In overseeing Abbott's Global Surveillance Program, a 26-year-old endeavor that started after the company developed the first U.S. Food and Drug Administration (FDA)-approved test for HIV in 1985, Mary Rodgers, principal scientist in diagnostics, has been instrumental in supporting Abbott's COVID-19 tests. "It was challenging and overwhelming at times. It typically takes 5 years for a test, and we launched 12 tests globally since the start of the pandemic. But every test result is a real person that gets to know about their health." For Rodgers, the stakes were especially high, given that her husband has a heart issue. "The fact that I was able to contribute to a test that saved my family is so meaningful to me and really surreal," she says.

Of course, the top employers' response to COVID-19 included more than tests and vaccines. The pandemic also propelled actions in corporate social responsibility, as these companies invested in more proactive and strategic approaches to diversity, equity, and inclusion (DEI); global climate change abatement initiatives; and social justice stances.

Insmed is taking an overall perspective and examining every corner of the company; it is now looking to reduce its carbon footprint and environmental impact. Recent initiatives include adding electric-vehicle charging stations in its parking lot and evaluating food-waste curtailment measures. "We have an eye toward being green," says Schaeffer. Abbott is investigating the environmental impacts of packaging to reduce plastic. Employees at Incyte are empowered to make a difference with autonomy and the ability to take ownership of a problem, which has led to improved sustainability measures. "We realized that we had a number of green initiatives in place, for example a green roof and tanks for holding and recycling water, which we had not spoken about formally and our employees were unaware of. In meetings, people would ask, 'Why aren't we greener?'" shares Swain. "So we decided to put a team in place to evolve and

cont.>

IMAGE: © PEDRONEVESDESIGN/SHUTTERSTOCK.COM



Great people and incredible work. That’s what it takes to break boundaries.

Join us.

It's among the most noble – and challenging – of all human endeavors. It's something everyone at Pfizer puts their heart and soul into every day.

Wherever they work with us and in whatever role, they're part of a global community – driven by a common purpose and striving for the same ambitious goals. They share the same uncompromising quality standards, and constantly seek out opportunities to move the industry on.

But most of all, each morning, when they head out to work, they know they're making a very real and positive difference in the world.

It's something that's a part of us all. If it's something you'd like to be part of too, find out more at pfizercareers.com.

pfizercareers.com

Breakthroughs that change patients' lives.

Pfizer

communicate around our sustainability efforts."

At Takeda, innovation extends beyond science to DEI, says Nurse, through an internal network of hundreds of volunteers who discuss ideas related to improving the culture to exceed DEI targets. "Takeda is a place where all employees' voices are heard. Innovation is a group effort, and I am very proud to be at our company, where DEI is a top priority and we have a fully written out plan with several pillars," she says.

DRIVING CHARACTERISTICS OF TOP EMPLOYERS



2021	2020
1. Innovative leader in the industry	1. Innovative leader in the industry
2. Treats employees with respect	2. Treats employees with respect
3. Is socially responsible	3. Is socially responsible
4. Has loyal employees	4. Has loyal employees
5. Easily adapts to change	5. Has work culture values that are aligned with employees' personal values

* Driving characteristics are listed in descending order of impact on overall employer rankings.
** Shaded background indicates the characteristics in common for the 2 years.

Empowering leadership

Leadership plays an important role in ensuring a company's success. At Insmed, leadership is a collective engagement. "We have no 'bosses,' no one who exclusively makes decisions without hearing the input from critical voices," says Perkins. "Everyone contributes and all voices are heard. We have supervisors who are more like mentors. In R&D, there's a collaborative spirit where everyone rolls up their sleeves." Anyone in the firm has access to the CEO and other senior leadership. This is similar to **Alnylam Pharmaceuticals**, a biopharmaceutical company based in Cambridge, Massachusetts (No. 3), where monthly All Company Meetings with the CEO are voluntarily attended by 90% of the company and there is an open invitation for feedback and ideas. "We set detailed goals with our colleagues to empower them to deliver the results," says Kelley Boucher, chief human resources officer. "Given the efforts we devote to setting our goals up front, colleagues have clarity on the role they play in achieving outcomes in a manner that is aligned with our core values."

Top companies invest in emerging leaders to sharpen their skills in uncovering novel opportunities, and to construct avenues for senior leadership to engage with staff across the enterprise in a two-way partnership. At Abbott, this comes in the form of cross-functional collaborations, in which employees from different departments have arenas to converse with each other, share ideas, and explore new realms for creativity; opportunities include internal conferences, monthly multidisciplinary scientific meetings, development courses, and events for employee resources groups, such as Women Leaders of Abbott. This approach also involves interaction with external partners, including academics, something that is available at other top employers. "There are all these unique routes for having new ideas bubble up and mechanisms to share those ideas," says Rodgers. At Insmed, global project teams provide a lever for innovation, where anyone can suggest ideas for advancement. "It allows exposure for junior people to develop [as leaders] in a safe manner," notes Perkins.

Leadership without authority is tethered to innovation. So top companies support efforts to enable their employees to be successful in coming up with and executing on the next big, potentially game-changing idea. "Abbott realizes that innovation doesn't happen by accident. You have to devote time and resources," says Rodgers. Those resources must include a fundamental mindset that is baked into the culture, which gives explicit permission to staff to pilot new endeavors. "We have empowered our staff to experiment or pivot to something new," says Schaeffer. "If you have an idea that is not in your functional area of expertise, we expect you to share and speak up, because we know we get a better result if people are open and transparent in sharing ideas." Echoes Maryanne Covington, senior director of Applied Technology Sciences at Incyte, about her experience at her company: "If one scientist has an idea, you can run it by the program. No matter how small the idea, your idea is heard."

Alnylam offers numerous leadership training and coaching programs that involve fireside chats with the CEO. The experience includes the chance to shadow colleagues across the company and complete a real-life case study about the firm's market penetration strategies. There is also a special program called Athena that is specifically for emerging women leaders.

Moderna (No. 7) also cultivates leadership throughout its organization. The biotech company headquartered in Cambridge, Massachusetts, was able to accelerate COVID-19 vaccine development in lightning speed because it invested in its leaders at all levels. "We work very hard on this," says Stephen Hoge, president of Moderna. "As a leader, I have a deep belief in the power of relationships. What we need to do is engage and communicate with a broad group of the organization, without silos, so people can solve the problem. We are at our best when we are in a mindset to invite passionate debate, act with healthy inquisition, and collaborate in pursuit of transforming medicine."

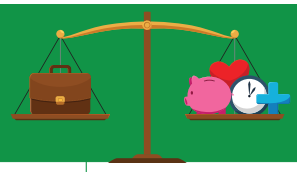
Supporting current and future STEM pipelines

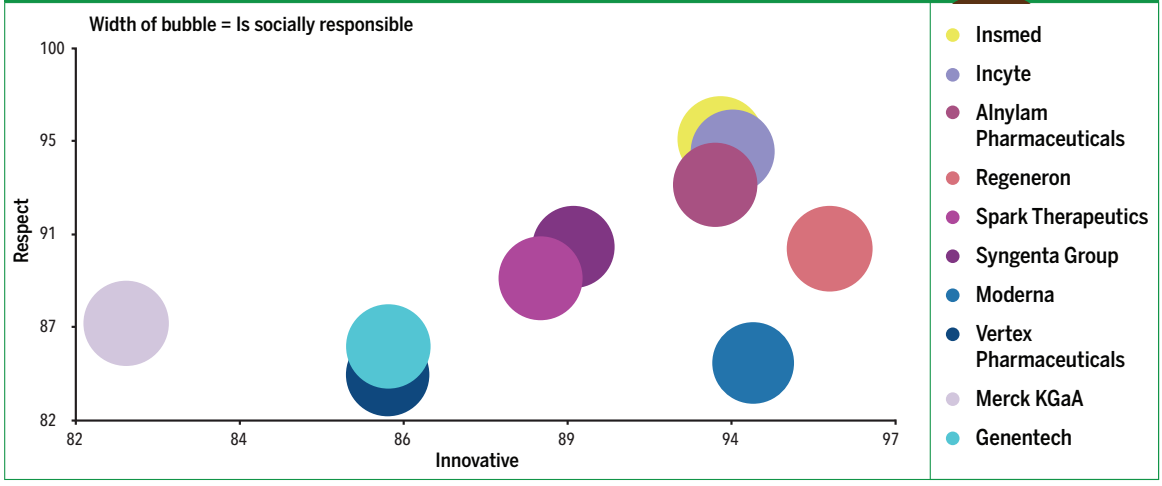
Employee development is something many scientists take seriously; they have indicated so in the Top Employers Survey. Abbott invests in its employees in numerous ways. Each scientist works through a personalized development plan, designed to give them guidance and resources to grow their career in the company, something that several top companies mirror. Abbott also provides various online learning opportunities in cont.>

IMAGE: © ASHATILOV/SHUTTERSTOCK.COM

the form of courses and even a mini-MBA program. "All of these development tools that help me be a leader at Abbott are available to all scientists here," says Rodgers. "It is part of Abbott's commitment to investing in voices in all areas of the company." Moreover, the performance of managers is assessed by how well they coach their direct reports and whether their team members are meeting their own career goals within the company. At Incyte, all employees—including fairly new Ph.D.s—are given an abundance of opportunities to lead programs and projects.

COMPARISON OF TOP TEN COMPANIES BY TOP THREE CHARACTERISTICS





Comparison of the top 10 companies on the basis of 3 of the top drivers (scored out of 100): *Is an innovative leader in the industry* (*x*-axis), *Treats employees with respect* (*y*-axis), and *Is socially responsible* (bubble width).

Such incentivizing of career development is seen at many of the top employers, such as Insmed, where employee development is supported through coaching, and upper management has regular discussions about how to better align talent within the organization in order to advance the company *and* the individual. Alnylam consistently develops 5-year plans that they share with all employees, which then feeds both annual corporate and individual objectives. "Our 5-year strategy serves as a North Star for colleagues to align on the mission, vision, and mechanisms by which we will achieve our long-term aspirational goals," says Boucher.

But it is not enough to improve the skills of current staff. A successful company must build and support a pipeline for up-and-coming scientific employees, too. This entails devoting resources to education. "Our commitment to scientific talent starts at the beginning: internships," says Kehr. "We are the number-one internship in health care and engineering. Once the interns join, we have systemic ways of developing their careers." Abbott's college internship is in its sixth decade and serves about 1,000 students each year. More than 150 students have participated in the company's high school internship since it started in 2012. Interestingly, 75% of Abbott's full-time engineers hired from the high school internship program are female. The program has become a strong diversity recruiting tool, notes an Abbott representative. At its New Jersey headquarters, Insmed has partnered with a local organization, Students 2 Science, which promotes science, technology, engineering, and mathematics (STEM) education in underserved New Jersey communities. Employees engage in mentoring and speaking about career paths and opportunities.

Nurturing health, flexibility, and fun

Some of the most influential organizations pursue a holistic approach in nurturing employees. "Living better, healthy lives is intrinsic to every level of the company. To me that is a really important combination," says Rodgers. That approach extends to the employees' families as well. Childcare offerings and flexible schedules for parents and caregivers are par for the course at these companies. "My boss told me 'you are the only mom for your kids.' That is the caring nature of my manager and the company," says Rodgers.

Indeed, work–life balance is universally important. "Insmed is a results-driven organization," says Schaeffer. "But there is always the expectation that if you are doing the job and delivering, it doesn't matter if you leave at 3 p.m. for a kid's soccer game or to help a parent. We don't manage our office by bed checks of who is in the office. It's about taking what you need to recharge and be with your family and be your best."

Flexibility is a key indicator of a top employer. As organizations shifted to remote work during the pandemic, top employers took full advantage of this, to reimagine a new way of doing business. Rather than see this as a deficit or damning scenario, the pharma and biotech companies were energized by the possibility of enabling their employees' success in a new environment. Not surprisingly, employees responded well to this. Boucher describes a colleague who had joined Alnylam 2 days before lockdown. An extrovert, the employee craved the chance to learn the company culture and contribute to its community. But with the company going fully remote, combined with a challenging home scenario—which included at-risk parents, homeschooled kids, a partner who was a frontline worker, and a recent pet who passed away—she knew this was going to be difficult. But what she didn't expect, says Boucher, was the company's caring and gracious response in accommodating and welcoming her. She told Boucher: "I cont.>

IMAGE: © KRISTNU/SHUTTERSTOCK.COM

Featured participants

Abbott
www.abbott.com
AbbVie
www.abbvie.com
Alnylam Pharmaceuticals
www.alnylam.com
Biocon Limited
www.biocon.com
Incyte
www.incyte.com
Insmed
insmed.com
Novartis
www.novartis.com
Novo Nordisk
www.novonordisk.com
Pfizer
www.pfizer.com
Regeneron
www.regeneron.com
Syngenta Group
www.syngenta.com
Takeda Pharmaceutical Company Limited
www.takeda.com
Vertex Pharmaceuticals
www.vrtx.com

knew I was joining extraordinary people when I came here, but the extraordinary support and kindness that leadership has provided me and my family during this pandemic is something I will never forget." Boucher smiles as she recalls this moment. "It really struck me how appreciative she was, and there wasn't an expectation she'd have this support," says Boucher. It became an emblem of pride—and of purpose—for Alnylam to continue to ensure that all staff have a similar experience.

There is one more ingredient needed in a fantastic workplace: fun! Companies engage employees in promoting relaxing activities, such as hosting virtual cooking classes and ninja workouts (led by the president), as is the case at Alnylam. Plump recorded funny carpool karaoke videos with his daughter that were broadcast across the company, jabbing bits of levity into an otherwise tumultuous era. "Keeping things fun and light while doing important and serious work is part of the secret sauce for how to live our core values, evolve our culture, and ultimately deliver important medicines for patients," says Boucher.

Valuing scientists—and science

In this year's survey, 78% of respondents declared they were not at all likely or not very likely to change jobs in the next year. It's not too much of a leap to infer that this is because the employers have shown them that they are valued.

The bottom line for top employers is culture, which is carefully assembled from values. Ultimately, the best organizations are those where individuals and their contributions are valued. They are collectives where the values of the whole mirror the values of the one. They are collaborations where the team values being able to be a part of something meaningful. "Culture is not just something that happens," says Plump. "You have to work for it with metrics that drive behavior. You can't just say, 'we will have a good culture' and not do anything. It will die off." Excellent employers make it part of their mission to engrain the culture into everything they do, creating a continuum that is regularly nurtured. "We talk about our values. We ask candidates in interviews about how they would operate within our values," says Schaeffer. "We measure performance not just by what is accomplished, but equally important, by how people achieve goals in a collegial and respectful way." Alnylam's core values—commitment to people, innovation and discovery, sense of urgency, open culture, and passion for excellence—flow into performance evaluations for team leaders and members, and are a continuous element of conversations about what the company and its staff should do next.



DEMOGRAPHICS

GENDER:
50% Male, 43% Female, 6% no response

EXPERIENCE:
72% have 10 or more years of work experience

HIGHEST DEGREE EARNED:
30% Doctorate, 33% Master's, 31% Bachelor's, 6% Other

COMPANY TYPE:
More than 9 out of 10 work in private industry: 28% Biotech, 42% Biopharma, 25% Pharma

NATURE OF WORK (TOP RESPONSES):
11% Basic Research, 15% Applied Research, 25% Development, 8% Production, 10% QA/QC or Regulatory Affairs, 9% Executive

GEOGRAPHY:
64% North America, 19% Europe, 11% Asia/Pacific Rim, 6% rest of world

But there is another thread common to top employers: Can your organization construct and nurture an ecosystem that affords its STEM-educated workforce the opportunity to do what they love, to leverage scientific problem-solving to help humanity, and in doing so, remind them why they got into science in the first place? Many top employers consciously offer this prospect and strive to make their organization a place that fosters employees' enthusiasm for science. "My love of science is being fed by my employment at Takeda," says Nurse. "One of the reasons I went into biostatistics is I wanted to use my math to help people. Here I get to do so. You get to see the fruits of your labor, science developed into innovative therapies to save patient lives. So, in my mind, it's a no-brainer why scientists would want to consider Takeda." Kathy Fernando, vice president and head of worldwide research, development and medical operations for **Pfizer** (No. 16), a New York City–based pharma firm, feels the same way about her employer. "There is a certain love of science here. You get to see that love and rigor translate into a benefit for the patients," she says. "It is so rewarding. It is powerful."

And when it comes to those working for top employers, that power and love means having the ability to have massive impact on humanity. As Abbott's Rodgers says, "I love that I get to do work that is meaningful in the world."

To read this article online and to view Top Employer profiles, visit ScienceCareers.org/TopEmployers.

Alaina G. Levine is a science writer, science careers consultant and coach, professional speaker, and author of Networking for Nerds *(Wiley, 2015).*

IMAGE: © SYUZANN/SHUTTERSTOCK.COM

MCGOVERN INSTITUTE

CALL FOR NOMINATIONS:

SCOLNICK PRIZE IN NEUROSCIENCE

The McGovern Institute for Brain Research is accepting nominations for the **18th annual Edward M. Scolnick Prize in Neuroscience**. The Prize recognizes an outstanding discovery or significant advance in the field of neuroscience. The prize is **$150,000**. The recipient presents a public lecture at MIT, hosted by the McGovern Institute and followed by a dinner in Spring 2022.

Nomination Deadline: **December 15, 2021**

Nomination procedures:
Candidates for the award must be nominated by individuals affiliated with universities, hospitals, medicals schools, or research institutes, with a background in neuroscience. Self-nomination is not permitted. Each nomination should include:

- A biosketch or CV of the nominee.
- A letter of nomination with a summary and analysis of the major contributions of the nominee to the field of neuroscience.
- Up to two representative reprints will be accepted.

Selection Procedure:

- Members of the selection committee and faculty affiliated with MIT are not eligible.
- Announcement of the award recipient will be made in January 2022.
- Recipient must attend all events to be awarded the prize.

Past Scolnick Prize Recipients:
2020: Joshua Sanes, Harvard University; **2019:** Richard Huganir, Johns Hopkins University; **2018:** David J. Anderson, HHMI, Caltech; **2017:** Catherine Dulac, HHMI, Harvard University; **2016:** Cornelia Bargmann, HHMI, The Rockefeller University; **2015:** Charles Gilbert, The Rockefeller University; **2014:** Huda Zoghbi, HHMI, Baylor University; **2013:** Thomas Jessell, HHMI, Columbia University; **2012:** Roger Nicoll, UCSF; **2011:** Bruce McEwen, The Rockefeller University; **2010:** Lily and Yuh-Nung Jan, UCSF; **2009:** Jeremy Nathans, Johns Hopkins University; **2008:** Michael Davis, Emory University; **2007:** David Julius, UCSF; **2006:** Michael Greenberg, Children's Hospital/HMS; **2005:** Judith Rapoport, NIH; **2004:** Masakazu Konishi, CalTech

Send nomination packet to:
gwolf@mit.edu or
Attn: **Scolnick Prize Nomination**
McGovern Institute for Brain Research
Massachusetts Institute of Technology
77 Massachusetts Avenue 46-3160
Cambridge, MA 02139

PURDUE UNIVERSITY

FACULTY OPENING IN INDUSTRIAL AND PHYSICAL PHARMACY

Assistant Professor or Associate Professor of Industrial and Physical Pharmacy

The College of Pharmacy at Purdue University is seeking applications to fill a faculty position in the Department of Industrial and Physical Pharmacy in the general area of **Pharmaceutical Biotechnology and/or Advanced Manufacturing.**

The position is for a full-time tenure-track Assistant Professor or Associate Professor. The faculty member is expected to establish an externally funded research program and will teach courses in the undergraduate, graduate and professional (PharmD) programs. It is expected that the individual will collaborate on various research activities within the department and in multidisciplinary research throughout the University. This is a nine-month (academic year) appointment.

The candidate must have expertise in fundamental and applied research related to the development, design, evaluation and manufacturing of biopharmaceutical products. Representative areas of expertise include, but are not limited to:

- Formulation-manufacturing interface (e.g., stability, analysis, and/or manufacturing) of peptide, protein, cell-based and/or nucleic acid-based therapeutics, vaccines, and/or immunotherapeutics.
- Advanced manufacturing approaches using novel technologies such as (but not limited to) continuous manufacturing, drying technology, aseptic processing, and 3D printing.

Established in 1869, Purdue is Indiana's land-grant university, a comprehensive educational and research institution which is a member of the prestigious American Association of Universities (AAU). The **Department of Industrial and Physical Pharmacy** encompasses multidisciplinary research and teaching in the areas of drug development, biotechnology, pharmacokinetics, drug delivery systems and manufacturing technologies.

Purdue University offers a thriving interdisciplinary environment for education and research. The university has a student body of over 49,000 undergraduate, graduate, and professional students, including 11 percent of underrepresented minority students and 18 percent international students.

Qualifications: The candidate must hold a PhD degree or equivalent in a relevant discipline such as pharmaceutical sciences, engineering, biochemistry, analytical chemistry and biophysics or closely related field. The candidate will have an outstanding record of scientific and technical accomplishment, an exceptional potential for world-class research and a commitment to both undergraduate and graduate education. Purdue University's Department of Industrial and Physical Pharmacy is committed to advancing diversity in all areas of faculty effort including scholarship, instruction, and engagement. Candidates should address at least one of these areas in a separate Diversity and Inclusion Statement, indicating their past experiences, current interests or activities, and/or future goals to promote a climate that values diversity and inclusion.

Application: For consideration, please upload a single PDF file that contains the following at **https://career8.successfactors.com/sfcareer/jobreqcareer?jobId=16500&company=purdueuniv**

- Candidate's curriculum vitae
- Letter of interest providing a summary of qualifications for the position
- Statement of research interests and goals
- Statement of teaching philosophy
- Statement of diversity and inclusion
- Names and contact information for three references

Incomplete applications will not be considered.

Review of applications will begin **November 15, 2021** and will continue until the position is filled. A background check will be required for employment in this position. The desired starting date is **August 15, 2022**, but is flexible and negotiable.

For further information concerning the search please contact Mary Ellen Hurt, Faculty Search Committee, Department of Industrial and Physical Pharmacy, College of Pharmacy, Purdue University, West Lafayette, IN 47907-2091 **mhurt@purdue.edu.**

Purdue University is an EOE/AA employer. Purdue University is committed to maintaining a community which recognizes and values the inherent worth and dignity of every person. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. All qualified applicants for employment will receive consideration without regard to race, religion, color, sex, national origin or ancestry, genetic information, marital status, parental status, sexual orientation, gender identity and expression, disability or status as a veteran.

ScienceCareers online @sciencecareers.org

SOMETIMES THE GRASS REALLY IS GREENER SOMEPLACE NEW.



Find your next job at ScienceCareers.org

There's scientific proof that when you're happy with what you do, you're better at what you do. Access career opportunities, see who's hiring and take advantage of our proprietary career-search tools. Get tailored job alerts, post your resume and manage your applications all in one place: sciencecareers.org

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

WAYNE STATE
School of Medicine

DEPARTMENT OF PHYSIOLOGY
Department Chair Position

With the rapid revitalization of Detroit, Wayne State University (WSU) is undergoing significant growth in its research enterprise. The Department of Physiology at WSU School of Medicine (SOM) (http://physiology.med.wayne.edu) invites applications for the Department Chair. We seek candidates that employ molecular, cellular or systems approaches to explore research interests in cardiovascular, metabolic, respiratory and other areas of physiology, pathophysiology and biophysics. WSU SOM is a state-of-the-art research environment, rated in the top third of all US Research Institutions by the Carnegie Foundation. The Department of Physiology has one of the most active research programs among the basic science departments at WSU-SOM and is presently ranked #35 out of ~120 Departments of Physiology in the USA. Start-up package and salary are highly competitive.

Candidates should hold a Ph.D., M.D. or equivalent from a relevant area. The Chair will have the vision and experience to lead a major urban academic Department of Physiology.

Please apply to **https://jobs.wayne.edu/applicants/jsp/shared/Welcome_css.jsp**, posting 045906, by uploading a curriculum vitae, a detailed future research plan, and names/contact information of three references. Please submit inquiries with a CV to **Physiologyfacultysearch@wayne.edu.** Review of applications will begin after **November 21, 2021** and continue until the positions are filled.

WSU is an Affirmative Action/Equal Opportunity Employer and encourages applications from women, people of color or other underrepresented backgrounds.

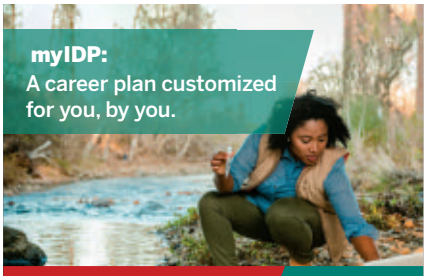
Deputy Director,
National Institute on Aging (NIA), NIH

The Deputy Director position at the National Institute on Aging (NIA) offers an exciting opportunity to serve as the second-in-command to the NIA Director in the scientific and managerial leadership of the Institute. The NIA, one of the 27 Institutes and Centers of NIH, leads broad scientific and research efforts to understand the nature of aging, aging processes, diseases and conditions associated with the nature of aging and extending healthy active years of life. The NIA is the primary federal agency responsible for research on Alzheimer's disease and related dementias.

Qualifications: a M.D. and/or Ph.D.; senior-level research experience and knowledge of research programs in one or more scientific areas related to aging research; and recognition and respect within their profession, both nationally and internationally, as individuals of outstanding scientific competence.

In addition, candidates must have demonstrated ability in: 1) planning, implementing, and analyzing program objectives and priorities, 2) managing financial and human resources, and 3) leading a research program involving extensive internal and external collaboration.

For full position details, please see the complete announcement located at **http://www.jobs.nih.gov/vacancies/executive.htm**. **Closes: Monday, November 29, 2021, Vacancy Contact: Brenda Fogel at SeniorRe@od.nih.gov**



There's only one Science

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.

myIDP Visit the website and start planning today!
myIDP.sciencecareers.org

Science Careers In partnership with:

UCSF University of California San Francisco · UIC · BURROUGHS WELLCOME FUND · University of Massachusetts Medical School · FASEB

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Confused about your next career move?

Download Free Career Advice Booklets!
ScienceCareers.org/booklets



Department of Ecology and Evolutionary Biology

TENURE-TRACK
ASSISTANT PROFESSOR
Ecology & Evolutionary Biology

The Department of Ecology & Evolutionary Biology at Tulane University invites applications for a tenure- track faculty position at the Assistant Professor level. We are seeking an evolutionary biologist with expertise in 'omics techniques and bioinformatics whose research will extend or complement existing strengths in the department. Candidates are required to have a Ph.D. in a related field at the time of application and postdoctoral research experience is preferred.

For details and to apply, please see **apply.interfolio.com/95618**

Tulane University is an Equal Opportunity/Affirmative Action/ADA Employer and encourages minority and female applicants to apply.

By **Katharina Maisel**

Your peers are your pillars

Phone in one hand, steaming cup of tea in the other, I opened Slack and started to scroll through my favorite channels. In #food_drink, pictures of enticing culinary creations offered inspiration (and the opportunity to drool a little). In #humblebrag, success stories brought a morning lift. Checking the app left me feeling connected to my community and ready to start the day with a positive mindset. This was totally new for me. In the past, I had avoided connecting with other scientists through online platforms—preferring face-to-face contact. But after starting as an assistant professor shortly before the onset of the COVID-19 pandemic, I had a new outlook.

Beginning a faculty position was surprisingly lonely. As I was facing the challenge of learning to manage a lab, I knew I would benefit from frank discussions with colleagues who were going through something similar. But I was the only new junior faculty member in my department. I wasn't sure where or how I would find the peer support I needed.

Years earlier, as an undergraduate student, I had benefited from the support of peers in my "living and learning" dorm community, which was composed of women in science and engineering. Being able to walk down the hall and get help on math problems I was having trouble with or code I just couldn't wrap my head around bolstered my confidence in the classroom.



"Checking the app left me feeling connected to my community and ready to start the day."

Later, when I was a graduate student and postdoc, I relied on my lab-mates. We spent long hours together not only doing research, but also during coffees, lunches, and happy hours, where we talked about confusing data, frustrations with assays not working, and successes to be celebrated. Sometimes knowing I was not the only one experiencing challenges was all I needed to keep going.

During that time, I also began to see how I could take steps to expand my peer mentoring opportunities. While having lunch with one of my Ph.D. classmates, she and I commiserated about missing the support from other women that we had enjoyed as undergrads. At some point the conversation turned to "why are we complaining and not acting?!" So we came up with a plan to connect with our peers and founded the Graduate Women's Empowerment Network. The group's board meetings and events deepened my connection with other women trainees at our institution and provided a forum for support when things got tough. Without this group, I would have been lost during the last 2 years of my Ph.D. program.

When I started my faculty position, I struggled at first because I was accustomed to having peer support but I didn't have anyone to connect with who was at the same level in my department. To try to fill the gap, I set up a regular informal meeting with my more senior women colleagues. I greatly valued their perspectives and support—but I still yearned to talk with researchers closer to my own career stage.

That's why I turned to New PI Slack, home of #food_drink, #humblebrag, and so much more. I had found the group through Twitter, which a colleague advised me to join shortly after I started my lab. After COVID-19 lockdowns began, I started to rely on it heavily, leaving it constantly running in the background on my phone and computer.

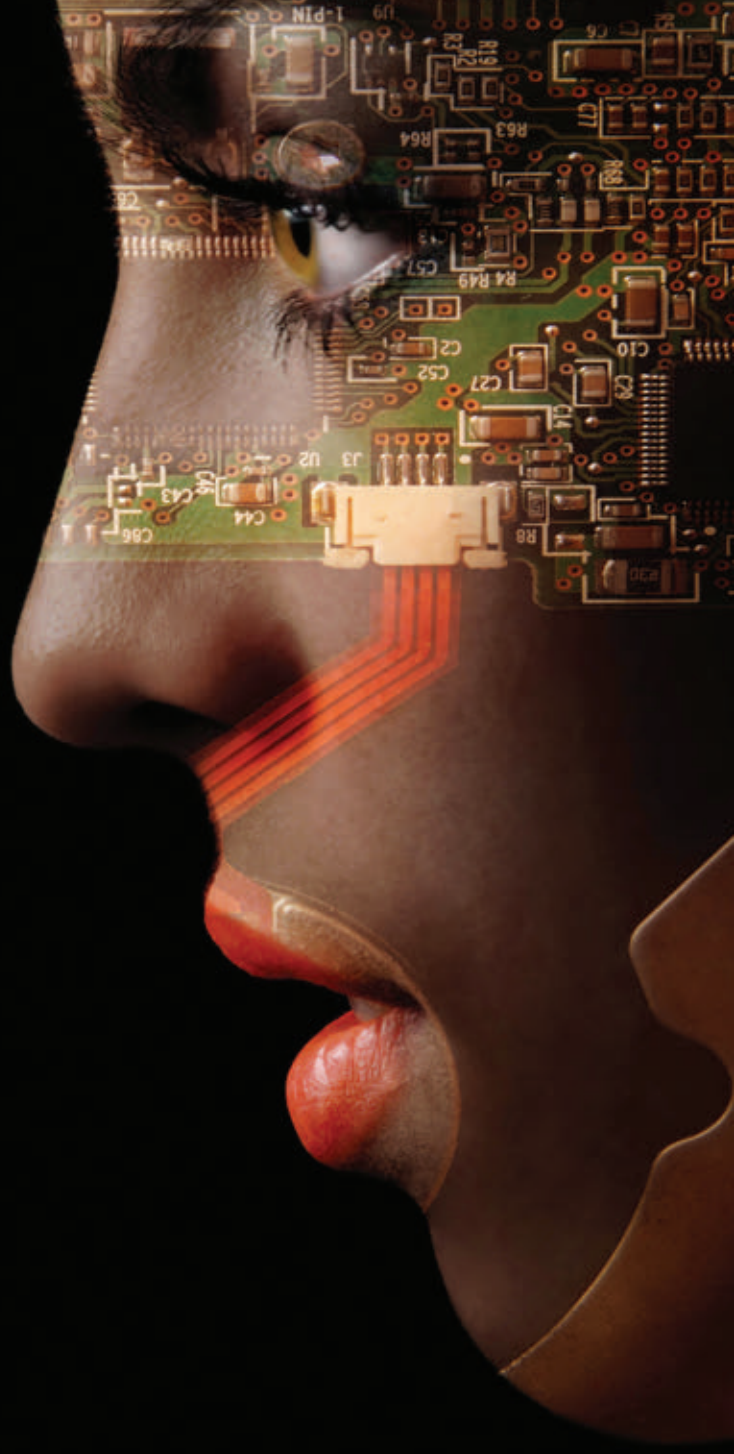
Connecting with others—whether swapping recipes, complaining about being tired of the same set of walks available near home, or sharing angst about how the pandemic would affect our careers—kept me sane, and was even invigorating and uplifting. I also learned from discussions about challenges at work, such as how to keep trainees motivated during a pandemic and ways to make research progress despite limited lab access. The support allayed any fears I had about wasting time online or struggling to connect with people I'd never met in person. The feeling of community was exactly what I and many others needed, and even led to new collaborations and friendships.

We often talk about mentoring from the top-down perspective. But I've learned that peer mentoring—whether virtual, in-person, or both—is just as important. Find your peer network pillars and lean into them. They will hold you up during the storm. ■

Katharina Maisel is an assistant professor at the University of Maryland, College Park. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Share **Your Robotics Research** with the **World**.



Shaping the future of robotics with high impact research!

As a multidisciplinary online-only journal, *Science Robotics* publishes original, peer-reviewed, research articles that advance the field of robotics. The journal provides a central forum for communication of new ideas, general principles, and original developments in research and applications of robotics for all environments.

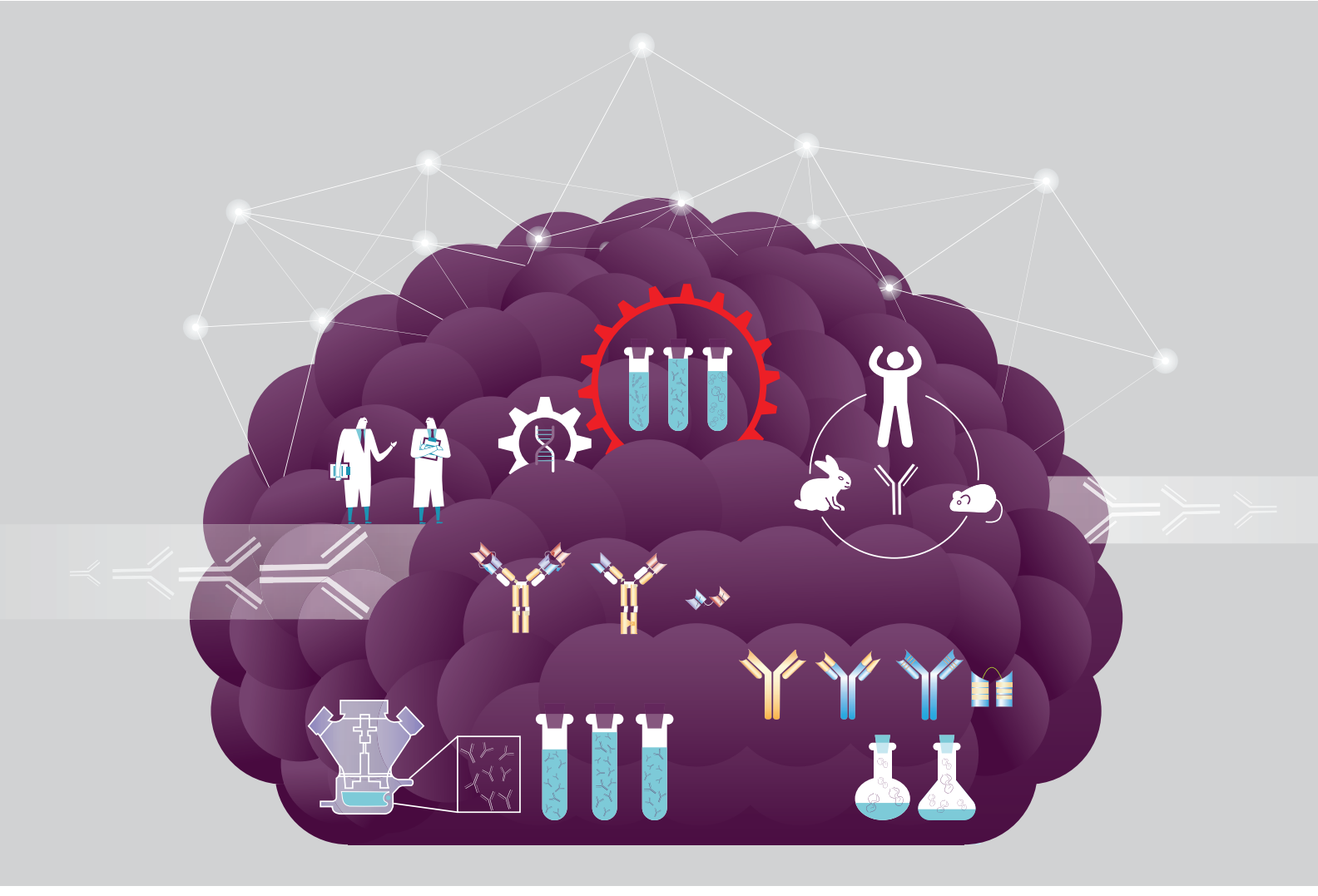
Submit your research today. Learn more at: **ScienceRobotics.org**

Twitter: @SciRobotics
Facebook: @ScienceRobotics

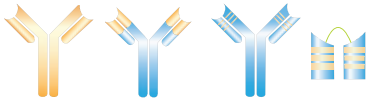
Science Robotics
AAAS

Streamlined Antibody Platform

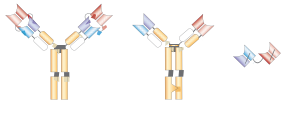
Antibody CRO for R&D and Manufacturing



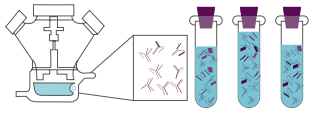
Antibody Humanization



Bispecific Antibodies



Stable Cell Line Generation



ProMab Biotechnologies offers nearly 20 years of experience to support antibody production, engineering and research and development. With over 5,000 hybridomas made, ProMab is committed to supporting antibody discovery.

All products are for research use only

Discover more | www.promab.com

ProMab Biotechnologies, Inc.

2600 Hilltop Dr, Building B, Suite C320, Richmond, CA 94806

1.866.339.0871 | info@promab.com

SCAN ME